



INSTITUTO TECNOLÓGICO SUPERIOR DE MISANTLA

“Identificación de patrones de comportamiento financiero utilizando el algoritmo DBSCAN”

T E S I S

PARA OBTENER EL GRADO DE MAESTRO EN
SISTEMAS COMPUTACIONALES

P R E S E N T A

ISC. ABIGAIL ECHEVERRIA CABRERA

ASESOR

DR. RAJESH ROSHAN BISWAL

CO ASESOR

DR. EDDY SÁNCHEZ DE LA CRUZ

Misantla, Veracruz

Agosto, 2018

RESUMEN

En el presente trabajo se presentan el análisis, preprocesamiento, procesamiento y aplicación del algoritmo de clusterización DBSCAN, el cuál fue aplicado a una base de datos la cual contienen información histórica de Agosto del 2004 a Diciembre del 2005, proveniente del Banco de México, dichos datos fueron recolectados del SPEI. El objetivo fundamental del presente trabajo es el preprocesamiento de los datos, la selección de los atributos necesarios para así poder definir cuáles serán utilizados para la implementación del algoritmo e identificar patrones y realizar la detección de anomalías. La metodología que se propone abarca como primera etapa el preprocesamiento de los datos, el primer paso de dicha etapa consiste en la unión de los archivos diarios dados originalmente por el Banco de México, obteniendo como final una estructura, como siguiente paso del preprocesamiento es la limpieza del ruido, el cual surge específicamente de los datos, puesto que inicialmente dentro de la estructura original cuenta con matrices de tamaño 130 que corresponde a los bancos, sin embargo muchos valores de ellos son cero, ya que no todos los bancos laboraban en el SPEI en el periodo de tiempo del cual se está trabajando, estos valores cero necesitan ser depurados para no causar incongruencias a la hora de tener los resultados finales, la limpieza antes mencionada fue trabajada en Matlab. Posteriormente se realizó una exportación de los datos que se tenían en Matlab, para poder darle continuidad en el lenguaje Python, en este efectuó un escalamiento de los datos para que los datos tuvieran una manera igualitaria, por ello se programó un *escalador* el cual convierte el contenido de la base de datos a valores entre 0 y 1. Como etapa final ya teniendo en cuenta los valores de y los archivos obtenidos por mes. Se procede con la aplicación del algoritmo DBSCAN. En la fase de experimentación se pueden apreciar las gráficas que muestran la evolución del comportamiento de cada uno de los bancos por día, formando grupos y haciendo la detección de las anomalías correspondientes a cada uno de ellos cuando encuentre un valor desbordado o insuficiente con respecto a los demás.

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

Estos resultados abren otras opciones para aplicaciones futuras en datos más recientes y así corroborar la detección de anomalías de bancos anteriores e incluso nuevos bancos.

ÍNDICE GENERAL

INTRODUCCIÓN.....	11
CAPÍTULO 1: GENERALIDADES	15
1.1 PLANTEAMIENTO DEL PROBLEMA	16
1.2 JUSTIFICACIÓN	18
1.3 OBJETIVO GENERAL.....	19
1.4 OBJETIVOS ESPECÍFICOS	19
1.5 HIPÓTESIS.....	20
1.6 PROPUESTA DE SOLUCIÓN	20
1.7 ALCANCES Y LIMITACIONES.....	21
1.7.1 ALCANCES	21
1.7.2 LIMITACIONES	21
1.8 ESTRUCTURA DE LA TESIS.....	22
CAPITULO 2: MARCO TEÓRICO	23
2.1 BANCO DE MÉXICO.....	24
2.2 SISTEMAS DE PAGOS	24
2.3 SPEI	25
2.4 ANOMALÍAS.....	26
2.5 DETECCIÓN DE ANOMALÍAS	26
2.6 TIPOS DE ANOMALÍAS	26
2.7 APRENDIZAJE AUTOMÁTICO (MACHINE LEARNING)	27
2.8 MINERÍA DE DATOS	28
2.9 CLUSTERIZACIÓN	29
2.10 MATLAB	30
2.11 PYTHON	31
2.12 SPYDER	32
2.13 WEKA	33
2.14 INDICADORES DE RED	34
2.14.1 Degree	34
2.14.2 Closeness	35
2.14.3 Betweenness	35
2.14.4 Affinity.....	36

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

2.15 ALGORITMO DE CLUSTERIZACIÓN: DBSCAN	37
CAPITULO 3: ESTADO DEL ARTE	38
3.1 CLASIFICACIÓN DE TÉCNICAS DE IDENTIFICACIÓN Y AGRUPACIÓN DE DATOS ANÓMALOS.....	39
3.2 CLUSTERING COMO HERRAMIENTA EN LA DETECCIÓN DE ANOMALÍAS	40
3.3 CLUSTERING BASADOS EN DENSIDAD.....	41
3.5 DBSCAN APLICADO EN LA DETECCIÓN DE ANOMALÍAS	42
CAPITULO 4: EXPERIMENTACIÓN	44
4.1 INTRODUCCIÓN	45
4.2 DATASET ORIGINAL.....	46
4.3 PRE PROCESAMIENTO DEL DATASET	49
4.3.1 Unión de base de datos diaria y eliminación de campos innecesarios	49
4.3.2 REDUCCIÓN DE BANCOS.....	50
4.3.3 OBTENCIÓN DE ARCHIVOS POR BANCO	51
4.3.4 GENERAR INDICADORES.....	52
4.4 PROCESAMIENTO DEL DATASET	55
4.4.1 Procesamiento con Python.....	55
4.4.2 Ejecución de DBSCAN por banco	57
4.5 ANÁLISIS DE BANCOS.....	60
4.6 COMPROBACIÓN DE RESULTADOS CON WEKA	64
CAPITULO 5: ANÁLISIS DE RESULTADOS.....	66
5.1 INTRODUCCIÓN	67
5.2 ANÁLISIS DE LOS BANCOS PARTICIPANTES.....	69
ANÁLISIS BANCO 1	69
ANÁLISIS BANCO 2	69
ANÁLISIS BANCO 3	71
ANÁLISIS BANCO 4	73
ANÁLISIS BANCO 5	74
ANÁLISIS BANCO 6	75
ANÁLISIS BANCO 7	76
ANÁLISIS BANCO 8	77
ANÁLISIS BANCO 9	79
ANÁLISIS BANCO 10	80
ANÁLISIS BANCO 11	82

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

ANÁLISIS BANCO 12	85
ANÁLISIS BANCO 13	86
ANÁLISIS BANCO 14	89
ANÁLISIS BANCO 15	90
ANÁLISIS BANCO 16	92
ANÁLISIS BANCO 17	94
ANÁLISIS BANCO 18	95
ANÁLISIS BANCO 19	97
ANÁLISIS BANCO 20	99
ANÁLISIS BANCO 21	100
ANÁLISIS BANCO 22	103
ANÁLISIS BANCO 23	105
ANÁLISIS BANCO 24	106
ANÁLISIS BANCO 25	108
ANÁLISIS BANCO 26	110
ANÁLISIS BANCO 27	113
ANÁLISIS BANCO 28	114
ANÁLISIS BANCO 29	116
ANÁLISIS BANCO 30	118
ANÁLISIS BANCO 33	120
ANÁLISIS BANCO 34	122
ANÁLISIS BANCO 35	124
ANÁLISIS BANCO 36	126
ANÁLISIS BANCO 37	128
ANÁLISIS BANCO 38.....	131
5.3 CONCLUSIONES Y TRABAJOS FUTUROS	133
ANEXO 1.....	134
Bibliografía	147

ÍNDICE DE FIGURAS

FIGURA 1. PROCESO KDD	12
FIGURA 2.1 TÉCNICAS Y ALGORITMOS EN LA MINERÍA DE DATOS	28
FIGURA 2.2. MATLAB (2018). MATHWORKS	30
FIGURA 2.3 PYTHON (2018). PYTHON.	31
FIGURA 2.4 SPYDER (2018). SPYDER	32
FIGURA 2.5 WEKA (2018). WEKA THE UNIVERSITY OF WAIKATO	33
FIGURA 4.1 ARCHIVOS MENSUALES DE MATLAB AGOSTO 2004- DICIEMBRE 2005	46
FIGURA 4.2 ESTRUCTURA INTERNA DE LOS ARCHIVOS DIARIOS SPEI	48
FIGURA 4.3 CÓDIGO UTILIZADO PARA LA UNIÓN DE BASE DE DATOS Y ELIMINACIÓN DE CAMPOS INNECESARIOS	49
FIGURA 4.4 CÓDIGO UTILIZADO PARA LA REDUCCIÓN DE BANCOS PARTICIPANTES DURANTE EL SPEI.....	50
FIGURA 4.5 CÓDIGO UTILIZADO PARA OBTENER LOS ARCHIVOS DE LOS BANCOS PARTICIPANTES.	51
FIGURA 4.6 ARCHIVOS DE LOS 38 BANCOS OBTENIDOS	51
FIGURA 4.6 MATRIZ DE ADYACENCIA	52
FIGURA 4.7 MATRIZ DE RECEPCIÓN DE PAGOS	52
FIGURA 4.8 MATRIZ W, NÚMERO DE RELACIONES DE LAS ENTIDADES POR DÍA	53
FIGURA 4.10 DEGREE DEL EJEMPLO	53
FIGURA 4.11 CÓDIGO PARA LA PROGRAMACIÓN DE INDICADORES.	54
FIGURA 4.12 CÓDIGO UTILIZADO PARA EL FORMATO DE INDICADORES.....	54
FIGURA 4.13 PROCESO DE CONVERSIÓN DE MATLAB - CSV - PYTHON	55
FIGURA 4.14 ESTRUCTURA INTERNA ARCHIVO MATLAB, BANCO 1	55
FIGURA 4.15 DATOS ORIGINALES CSV- BANCO1.....	56
FIGURA 4.16 DATOS ESCALADOS PRIMERA PARTE, BANCO 1.....	57
FIGURA 4.17 ALGORITMO PYTHON CON ESCALADO DE DATOS Y EJECUCIÓN DEL ALGORITMO DBSCAN	59
FIGURA 4.18 ANÁLISIS BANCO 10. CLASIFICACIÓN DE DÍAS CON SUS CLÚSTER Y SUS RESPECTIVAS	60
FIGURA 4.19. GRÁFICA DE RESULTADOS ANOMALÍAS OBTENIDAS. BANCO 1	61
FIGURA 4.20 COMPROBACIÓN DE RESULTADOS EN WEKA. IDENTIFICACIÓN DE DÍAS CON RUIDO (ANÓMALOS).....	65

ÍNDICE DE TABLAS

TABLA 4.1 PARÁMETROS UTILIZADOS EN EL ALGORITMO DBSCAN PARA LA FASE DE EXPERIMENTACIÓN	58
TABLA 4.2 EVALUACIÓN DE RESULTADOS DE ACUERDO A SUS PARÁMETROS.....	58
TABLA 4.3 ANÁLISIS BANCO 10. IDENTIFICACIÓN DE ANOMALÍAS POR INDICADOR DEPENDIENDO SU MEDIA Y DESVIACIÓN ESTÁNDAR	62
TABLA 4.4 MEDIA Y DESVIACIÓN ESTÁNDAR OBTENIDA POR CADA CLUSTER. COMPROBACIÓN DE DÍAS ANÓMALOS.	63
TABLA 5.1 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 1.....	68
TABLA 5.2 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 1.....	69
TABLA 5.3 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 2.....	69
TABLA 5.4 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 2.....	70
TABLA 5.5 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 3.....	71
TABLA 5.6 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 3.....	72
TABLA 5.7 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 4.....	73
TABLA 5.8 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 4.....	73
TABLA 5.9 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 5.....	74
TABLA 5.10 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 5.....	74
TABLA 5.11 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 6.....	75
TABLA 5.12 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 6.....	75
TABLA 5.13 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 7.....	76
TABLA 5.14 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 7.....	77
TABLA 5.15 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 8.....	78
TABLA 5.16 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 8.....	78
TABLA 5.17 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 9.....	79
TABLA 5.18 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 9.....	80
TABLA 5.19 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 10	81
TABLA 5.20 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 10.....	82

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

TABLA 5.21 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 11	83
TABLA 5.22 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 11.....	84
TABLA 5.23 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 12	85
TABLA 5.24 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 12.....	86
TABLA 5.25 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 13	87
TABLA 5.26 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 13.....	88
TABLA 5.27 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 14	89
TABLA 5.28 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 14.....	90
TABLA 5.29 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 15	91
TABLA 5.30 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 15.....	91
TABLA 5.31 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 16	93
TABLA 5.32 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 16.....	93
TABLA 5.33 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 17	94
TABLA 5.34 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 17.....	95
TABLA 5.35 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 18	95
TABLA 5.36 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 18.....	96
TABLA 5.37 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 19	97
TABLA 5.38 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 19.....	98
TABLA 5.39 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 20	99
TABLA 5.40 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 20.....	100
TABLA 5.41 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 21	101
TABLA 5.42 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 21.....	102
TABLA 5.43 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 22	103
TABLA 5.44 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 22.....	104
TABLA 5.45 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 23	105
TABLA 5.46 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 23.....	106

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

TABLA 5.47 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 24	106
TABLA 5.48 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 24.....	107
TABLA 5.49 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 25	109
TABLA 5.50 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 25.....	109
TABLA 5.51 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 26	111
TABLA 5.52 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 26.....	112
TABLA 5.53 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 27	113
TABLA 5.54 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 27.....	113
TABLA 5.55 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 28	115
TABLA 5.56 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 28.....	115
TABLA 5.57 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 29	116
TABLA 5.58 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 29.....	117
TABLA 5.59 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 30	119
TABLA 5.60 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 30.....	119
TABLA 5.61 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 31	120
TABLA 5.62 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 31.....	121
TABLA 5.63 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 34	123
TABLA 5.64 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 34.....	123
TABLA 5.65 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 35	124
TABLA 5.66 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 35.....	125
TABLA 5.67 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 36	127
TABLA 5.68 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 36.....	127
TABLA 5.69 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 37	129
TABLA 5.70 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 37.....	130
TABLA 5.71 ANÁLISIS DE ANOMALÍAS DETECTADAS EN BANCO 38	132
TABLA 5.72 ANÁLISIS ESTADÍSTICO DE LA MEDIA Y DESVIACIÓN ESTÁNDAR DE LOS INDICADORES DE RED EN BANCO 38.....	132

INTRODUCCIÓN

Uno de los factores principales que surgen al analizar datos es la existencia de grandes volúmenes de estos, ya que, dependiendo el sector que se analice, en ocasiones sobrepasan la capacidad de recolectar, almacenar y comprender dichos datos. Para este estudio es utilizada la minería de datos, la cual nos ayuda a extraer el conocimiento a partir de un conjunto de información que las empresas han generado. Dicho conocimiento adquirido genera distinto tipo de aplicaciones las cuales van desde la prevención de fraudes y abusos, análisis de fidelización, cross-selling, hasta el análisis de datos que tienen un comportamiento distinto al resto, a lo cual denominamos Detección de Anomalías (DA).

Para poder realizar la detección de anomalías es necesario aplicar para su análisis un algoritmo de minería de datos basándose de igual forma en una técnica de clustering, esta técnica nos será de utilidad pues nos posibilita el reconocimiento de grupos de datos cuyo comportamiento es diferente al resto.

El clustering identifica clúster o regiones densamente pobladas tomando en cuenta previamente alguna medida de distancia en un conjunto de datos multidimensional como es el caso.

El análisis de clustering es basado en distintos tipos de métodos como son: jerárquicos, de partición, basados en densidad, basados en cuadrículas, restricciones y escalables. Como en todo análisis existen algunas debilidades de esta técnica, las cuales consisten en una mala elección de las métricas, combinar variables, entre otras. Dentro de los algoritmos basados en clustering se tienen el DBSCAN, Sequence, Kohonen, TwoStep.

Los algoritmos de agrupamiento que se emplean son un método de aprendizaje no supervisado puesto que se lleva a cabo sobre un conjunto de datos de entrada del cual no se tiene información, por lo tanto el sistema debe ser capaz de reconocer patrones para así poder etiquetar las nuevas entradas.

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

Existen 3 metodologías predominantes para el proceso empleado en la minería de datos son: KDD, CRISP-DM y SEMMA. Como parte del proceso de desarrollo en el estudio del dataset del banco de México se ha optado por aplicación de la metodología KDD que hace referencia a su nombre en inglés Knowledge Discovery in Databases o mejor conocida como Descubrimiento de conocimiento en Bases de datos, dicha metodología identifica patrones potencialmente útiles y entendibles. El proceso que sigue la metodología KDD se muestra en la siguiente Figura 1.

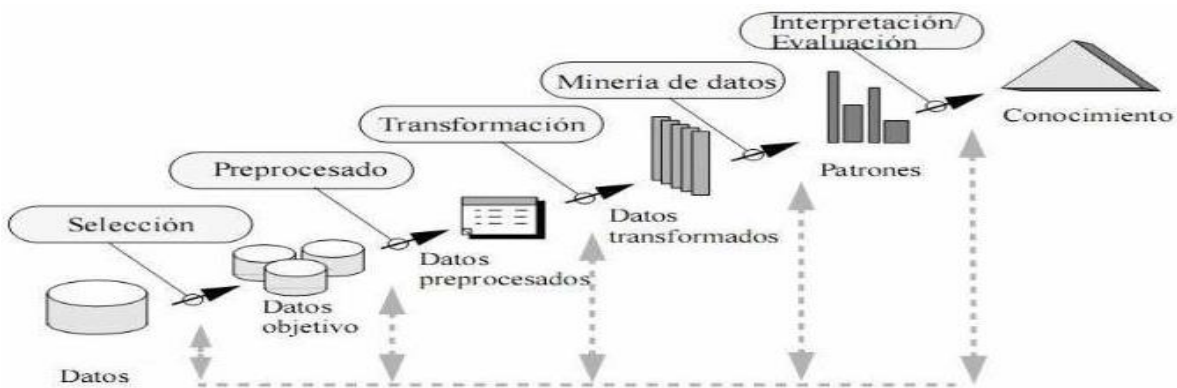


Figura 1. Proceso KDD

Dicha metodología propone como primera etapa seleccionar y extraer los datos con los que se trabajarán, como segunda etapa el pre procesamiento el cual consiste básicamente en la preparación y limpieza de los datos extraídos hasta dejarlos de una forma manejable, la tercera etapa es la transformación que es el tratamiento preliminar de los datos, como la generación de nuevas variables a partir de las que ya existen, en la cuarta etapa nos encontramos como *Data Mining* la cual es la fase del modelamiento, el objetivo es extraer patrones desconocidos, nuevos y potencialmente útiles que se encuentran contenidos u ocultos en los datos, como última etapa se tiene la interpretación y evaluación que tiene como objetivo identificar los patrones que se obtuvieron basándose en algunas medidas y para finalizar realizar una evaluación de los resultados obtenidos.

En muchos dominios incluyendo la detección de anomalías los datos se recopilan en forma de secuencias o series de tiempo para su análisis, estas series de tiempo

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

son referidas a datos estadísticos que se recopilan, observan o registran en intervalos de tiempo regulares (diario, semanal, semestral anual, entre otros. El termino serie de tiempo es aplicado a datos registrados en forma periódica. En este caso de estudio los datos financieros emitidos por el SPEI se encuentran en registros diarios que corresponden a cada uno de los bancos activos, en total los días que se analizan son 355 días trabajados dentro de las fechas anteriormente mencionadas. El objetivo de analizar bajo una serie de tiempo tiene diversos objetivos dentro de los que debemos destacar los siguientes:

- Determinar si se presentan ciertos patrones o pautas no aleatorias.
- Aislar y estudiar los componentes a fin de proporcionar claves para el futuro.
- Posibilitar el pronóstico de movimientos futuros así como otros aspectos relacionados.

Distinguir entre el ruido en datos normales y anomalías auténticas es difícil y la clasificación errónea conduce a un alto índice de falsos positivos o falsos negativos.

El Banco de México es el banco central del Estado Mexicano, autónomo en sus funciones y administración, tiene como finalidad proveer a la economía del país, teniendo como objetivo prioritario procurar la estabilidad del poder adquisitivo de la moneda y propiciar el buen funcionamiento de los sistemas de pago.

Los sistemas de pago son medios a través de los cuales se transfieren fondos entre bancos, son un requisito fundamental para mantener y promover la estabilidad financiera asegurando la circulación del dinero. Estos sistemas de pago han sido catalogados en dos grandes grupos:

Sistemas de pago de alto valor que son aquellos pagos urgentes por montos altos, algunos de estos pertenecen a pagos de empresas y liquidación de mercados financieros. Un ejemplo de estos sistemas de pago son: SPEI, Fedwire, CHIPS, SIC, CHAPS. Por otra parte tenemos a los sistemas de pago de bajo valor que se refieren a los pagos menos urgentes por montos menores, tales como: pagos entre particulares, comercios y nóminas. Algunos ejemplos de estos sistemas de pagos son: cheques, tarjetas, transferencias electrónicas, y domiciliaciones.

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

El sistema de pago empleado por el Banco de México es el SPEI perteneciente y operado por él mismo, el cuál consiste en un sistema de transferencias electrónicas de fondos, el objetivo principal de este sistema es facilitar los pagos entre las instituciones financieras y sus participantes, prácticamente en tiempo real, las 24 horas del día, todos los días del año.

En el siguiente trabajo se encarga de analizar la red de pagos del SPEI en busca de anomalías en determinadas series de tiempo haciendo uso de técnicas de Machine Learning.

CAPÍTULO 1

GENERALIDADES

1.1 PLANTEAMIENTO DEL PROBLEMA

Una anomalía según (Giménez García, 2009), se define como la existencia de una discrepancia de una regla o de un uso, para poder detectar dicha discrepancia, se tiene que definir primero lo que se considera como un comportamiento normal de un sistema, y posteriormente se realizará una clasificación sospechosa o intrusiva a aquellos comportamientos que se desvíen de lo que se considera como normal.

La detección de anomalías es el proceso de monitorear las situaciones que ocurren en un sistema o red de cómputo y examinarlos en busca de signos de algún tipo de datos anómalos. Según (Qayyum, Islam, & Jamil, 2005), la detección de anomalías se efectúa detectando cambios en los patrones de utilización de los recursos o cambios en el comportamiento del sistema, se puede llevar a cabo construyendo un modelo estadístico con métricas derivadas de la operación normal del sistema, y marcar como intrusivo cualquier evento cuya métrica observada presente una desviación estadística significativa del modelo estadístico.

En otras palabras, una detección de anomalías usa un modelo “normal” del comportamiento de la red para compararlo con el comportamiento actual observado. Cualquier comportamiento que muestre un cambio significativo con respecto al modelo, se considera como una anomalía.

El problema que se plantea en esta investigación se enfoca en el análisis de datos y así detectar posibles anomalías en un dataset el cual contiene los datos financieros proporcionados por el Banco de México los cuales datan desde Agosto del 2004 hasta Diciembre 2005, teniendo éstos como un conjunto de entrenamiento que nos ayudará en la búsqueda de patrones de comportamiento, de forma posterior se irá incrementando el conjunto para así poder obtener mejores resultados.

El análisis del dataset en cuestión además de permitir conocer el comportamiento de cada uno de los bancos que son estudiados, busca reducir de

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

manera considerable en un futuro el tiempo de búsqueda de posibles anomalías que pudiesen resultar susceptibles de una posible acción fraudulenta.

El área del conocimiento en minería de datos ideal para trabajar con la detección de anomalías es el machine learning. Según (Jiawei Han, Micheline Kamber 2001) citado por (González Bernal, 2017, pág. 7) define a la minería de datos como el proceso de descubrir conocimiento interesante de grandes cantidades de datos almacenadas en bases de datos, datawarehouses u otro repositorio de información. Uno de estos conocimientos es la detección de anomalías.

1.2 JUSTIFICACIÓN

El aporte del presente trabajo es analizar el comportamiento de cada banco y así ayudar al Banco de México a entender de mejor manera la conducta actual de cada uno de los bancos, de igual manera se procede a identificar más fácilmente cuando un participante no se comporte de manera habitual, es decir, que tenga un comportamiento anómalo.

Es por eso que es necesario hacer uso de un algoritmo de minería de datos, que nos ayude a identificar dichos comportamientos de cada banco y así detectar anomalías.

Mediante la aplicación de minería de datos (MD) sobre el dataset la meta es encontrar objetos anómalos, o lo que es igual, objetos que sean diferentes a los demás también conocidos como *outliers*, puesto que se refieren a detectar desviaciones de objetos anómalos que tienen atributos con una desviación significativa con respecto a los valores típicos esperados, éstas anomalías obtenidas la mayoría de los casos son tratadas como ruido o errores en operaciones tales como el clustering.

Una vez obtenidos estos datos anómalos se debe encontrar una manera de identificarlos y agruparlos, es aquí donde se pueden clasificar en técnicas basadas en modelos, proximidad o densidad expuestas más adelante. Esta técnica de agrupación con base en un criterio que por lo general son distancia o similitud es conocida como *clustering*.

1.3 OBJETIVO GENERAL

Analizar los indicadores de red obtenidos del SPEI para así poder ejecutar el algoritmo DBSCAN el cual permitirá identificar los comportamientos de los bancos y con base en esto poder detectar anomalías (outliers).

1.4 OBJETIVOS ESPECÍFICOS

- Seleccionar y programar los indicadores para el procesamiento de los datos.
- Realizar una estructura de base de datos proveniente del pre-procesamiento.
- Realizar el escalamiento de los datos para su procesamiento, obteniendo una estructura final.
- Implementar un algoritmo de *Data Mining* sobre la estructura generada.
- Analizar y validar resultados.

1.5 HIPÓTESIS

H1: Realizando la unión de los archivos dados para la obtención de un dataset final, haciendo la selección correspondiente de los atributos mediante los indicadores de red, así como aplicando el algoritmo adecuado, es posible determinar los días en los que se detectó un comportamiento diferente según sea cada caso con respecto a los bancos.

H2: Mediante la unión de los archivos dados para la obtención de un dataset final, haciendo la selección correspondiente de los atributos mediante los indicadores de red, es posible determinar los días anómalos y el clúster en el que se encuentran, posteriormente determinar estadísticamente el indicador por el cual dicho día es anómalo, así como tener el conocimiento del comportamiento de cada uno de los bancos participantes.

Para conocer la aprobación o no de dichas hipótesis anteriormente mencionadas es necesario comparar los resultados obtenidos con otro dataset con un mayor número de días, o bien compararlos con lo que se describe en el estado del arte teniendo en cuenta que las implementaciones del algoritmo aplicado a una empresa no son enfocadas al sector financiero.

1.6 PROPUESTA DE SOLUCIÓN

Como propuesta de solución, se expone analizar previamente los datos obtenidos de la red de pagos SPEI en el periodo de tiempo establecido por el banco desde su inicio en agosto 2004 hasta Diciembre 2005, con esto realizar el pre-procesamiento de los datos y posteriormente separar en archivos distintos por cada banco, los cuales a su vez, se separarán en archivos mensuales, que serán procesados con el algoritmo DBSCAN en búsqueda de anomalías. Todo lo anterior se realiza en las siguientes fases:

Fase 1: Unir y pre procesar los datos en Matlab, limpiarlos y quitar el ruido existente de los bancos que no iniciaron con el SPEI.

Fase 2: Ejecutar los algoritmos los cuales contienen los indicadores de red previamente establecidos por el Banco de México.

Fase 3: Separar los archivos por cada banco para obtener una estructura final.

Fase 4: Escalar los datos, para estandarizar valores y tener resultados igualitarios.

Fase 5: Aplicar el algoritmo de clustering en Python.

Fase 6: Analizar los resultados obtenidos en las gráficas para identificar si surgieron datos anómalos en cada uno de los bancos.

1.7 ALCANCES Y LIMITACIONES

1.7.1 ALCANCES

Implementar el algoritmo DBSCAN sobre los indicadores programados haciendo uso de la base de datos de montos y envíos de Agosto 2004 a Diciembre 2005, en los 38 bancos participantes en el inicio del SPEI obteniendo finalmente un análisis exclusivo de cada uno de los bancos acentuando su comportamiento, así como en la obtención de datos anómalos o comportamientos irregulares, haciendo una representación con gráficas donde se muestre dicho análisis.

1.7.2 LIMITACIONES

Se trabajará actualmente con los datos obtenidos de la red de pagos SPEI, de Agosto 2004 a Diciembre 2005 para su análisis, además de no conocer el nombre de cada uno de los participantes, todo lo anterior por motivos de políticas de privacidad del banco de México. Se utiliza el algoritmo *DBSCAN* el cual se encuentra en Python utilizando la biblioteca *scikit-learn*.

1.8 ESTRUCTURA DE LA TESIS

En el capítulo I se describe de manera particular la problemática existente, la manera de brindar solución ha dicho problema, justificando cada uno de los puntos expuestos del porque y como ayudaría al banco de México le análisis y la aplicación del algoritmo DBSCAN sobre los datos. Se considera el objetivo general y los objetivos específicos a perseguir para lograr la meta propuesta, así como los alcances de nuestro proyecto y las limitaciones que pueden surgir durante el proceso de nuestra investigación y aplicación.

En el capítulo II define los conceptos expuestos durante el presente trabajo, es decir interviene toda la teoría que representa nuestra investigación con la finalidad de conocer de manera pertinente todo lo que se abordará en el transcurso del trabajo.

Durante el capítulo III se realiza una recopilación de los trabajos anteriores, analizando y elaborando una comparación de los resultados obtenidos con el algoritmo aplicado en dichos trabajos y el que se muestra a continuación.

El capítulo IV denominado Experimentación, de primera instancia explica el contenido original del dataset y describe poco a poco cómo fue el pre procesamiento de esos datos y separación en series de tiempos, para finalmente poder ejecutar el algoritmo DBSCAN en búsqueda de la detección de anomalías.

El capítulo V analiza los resultados obtenidos en la ejecución del algoritmo DBSCAN, muestra más de 38 gráficas sobre el comportamiento de cada uno de los bancos participantes con respecto a los 355 que trabajó cada uno de ellos.

El capítulo VI muestra con base en los resultados mostrados el capítulo anterior, las conclusiones pertinentes respecto al comportamiento de los participantes del SPEI, así como los trabajos futuros que podrían realizarse a partir de este trabajo de tesis, y los beneficios que este documento podría aportar hacia ellos.

CAPITULO 2

MARCO TEÓRICO

2.1 BANCO DE MÉXICO

El Banco de México es el banco central del Estado Mexicano, autónomo en sus funciones y administración, su finalidad es proveer a la economía del país de moneda nacional y como objetivo prioritario es procurar la estabilidad del poder adquisitivo de dicha moneda, además de promover el sano desarrollo del sistema financiero y propiciar el buen funcionamiento de los sistemas de pago.

Banco de México considera que es de suma importancia mejorar la comprensión pública de qué es y qué hace la banca central en nuestro país, en particular respecto de sus acciones encaminadas a mantener la estabilidad de precios, procurar el sano desarrollo del sistema financiero, garantizar el buen funcionamiento de los sistemas de pago y proveer un medio de intercambio seguro y confiable para que las personas puedan realizar sus transacciones económicas. (Banco de México, 2018)

2.2 SISTEMAS DE PAGOS

Los sistemas de pagos están constituidos por un conjunto de instrumentos, procedimientos y normas para transferir recursos financieros entre sus participantes. Dichos sistemas son indispensables para que el sistema financiero funcione eficientemente. Algunos de ellos son especialmente críticos ya que si su diseño no es adecuado, pueden magnificar la transmisión de problemas de liquidez de un participante a los demás y perturbar la estabilidad del sistema financiero. Por estas razones, uno de los objetivos del Banco de México es propiciar el buen funcionamiento de los sistemas de pago del país. (Banco de México, 2017)

Los sistemas de pago son los medios a través de los cuales se transfieren fondos entre bancos, son el principal canal por el que se pueden transmitir perturbaciones entre los sistemas y los mercados financieros nacionales e internacionales. Los sistemas de pago sólidos son, por consiguiente, un requisito

fundamental para mantener y promover la estabilidad financiera, asegurando la circulación del dinero. (Internacionales, 2001, pág. 6)

Los sistemas de pago han sido catalogados en dos grandes grupos:

Sistemas de pago de alto valor

- Pagos urgentes por montos altos
- Pagos de empresas, liquidación de mercados financieros.
- Ejemplos: SPEI, Fedwire, CHIPS, SIC, CHAPS.

Sistemas de pago de bajo valor

- Pagos menos urgentes por montos bajos
 - Pagos entre particulares, comercios, nóminas.
 - Ejemplos: Cheques, tarjetas, transferencias electrónicas, domiciliaciones.
- (México, CEMLA, 2013)

2.3 SPEI

El SPEI es un sistema de transferencias electrónicas de fondos que pertenece al Banco de México y es operado por el mismo. . El SPEI permite a sus participantes realizar pagos en pesos mexicanos por cuenta propia y a nombre de sus cuentahabientes, prácticamente en tiempo real, las 24 horas del día, todos los días del año. Este sistema se desarrolló con el objetivo de:

- Facilitar los pagos entre las instituciones financieras
- Habilitar las instituciones financieras para ofrecer a la población servicios de pago al menudeo de forma segura y eficiente.
- Permitir a sus participantes realizar pagos en pesos mexicanos por cuenta propia y a nombre de sus cuentahabientes, prácticamente en tiempo real, las 24 horas del día, todos los días del año. (México, Banxico, 2016)

2.4 ANOMALÍAS

Las anomalías se basan en la observación de desviaciones de los patrones de uso normales en el sistema. Se las detecta construyendo previamente un perfil del sistema a monitorizar y posteriormente estudiando las desviaciones que se produzcan con respecto a este perfil. (informática, 2018)

2.5 DETECCIÓN DE ANOMALÍAS

(Jiménez, 2009), define que la detección de anomalías se basa en la suposición de que el comportamiento que tienen los usuarios y la red es lo suficientemente regular como para sacar un patrón, de forma que cualquier desviación significativa pueda ser considerada como una evidencia de una intrusión en el sistema.

2.6 TIPOS DE ANOMALÍAS

Para su estudio las anomalías se dividen en dos clases:

Basadas en similitud

Puntuales: Son el tipo más simple de una anomalía, se describe mediante la observación de características y atributos, la anomalía puntual solo llega a afectar a alguno de ellos, tratando así a las características de manera independiente.

Contexto: Es una ampliación de una anomalía de tipo puntual, ya que aquí se tienen en cuenta varias características a la vez, de esta forma las características tienen dependencia.

Colectivas: Se detectan al observar una colección de observaciones. Las anomalías colectivas se manifiestan al analizar un conjunto de observaciones. La

observación en sí misma no es anómala, pero un grupo de tales observaciones que ocurren juntas constituyen una anomalía. A menudo, ese grupo se forma al considerar la dimensión del tiempo.

Reglas Raras

Ambas clases comparten el tipo de anomalía contextual. Un elemento es llamado anomalía contextual, cuando su comportamiento se considera anómalo dependiendo el contexto (escenario) en el que se encuentra. Para determinar esto son necesarios dos atributos: Contextuales (determina el escenario o vecindario donde se encuentra el elemento) y de comportamiento (las características de ese elemento).

2.7 APRENDIZAJE AUTOMÁTICO (MACHINE LEARNING)

Es una disciplina científica del ámbito de la Inteligencia Artificial que crea sistemas que aprenden automáticamente. Aprender en este contexto quiere decir identificar patrones complejos en millones de datos.

La máquina que realmente aprende es un algoritmo que revisa los datos y es capaz de predecir comportamientos futuros. Automáticamente, también en este contexto, implica que estos sistemas se mejoran de forma autónoma con el tiempo, sin intervención humana. (González, 2014)

En el aprendizaje automático, se adquiere el conocimiento a través de analizar datos y manipularlos, usando estrategias basadas en teorías matemáticas. (Gómez, 2016)

2.8 MINERÍA DE DATOS

La minería de datos tiene como propósito la identificación de un conocimiento obtenido a partir de las bases de datos que aporten hacia la toma de decisión, es la exploración y análisis de grandes volúmenes de datos con el objeto de encontrar patrones y reglas significativas (conocimiento). Extracción de patrones y modelos interesantes, potencialmente útiles y datos en base de datos de gran tamaño. (Aguilar, 2018). En el siguiente esquema de la Figura 2.1 se muestran los tipos de técnicas utilizadas en la MD.

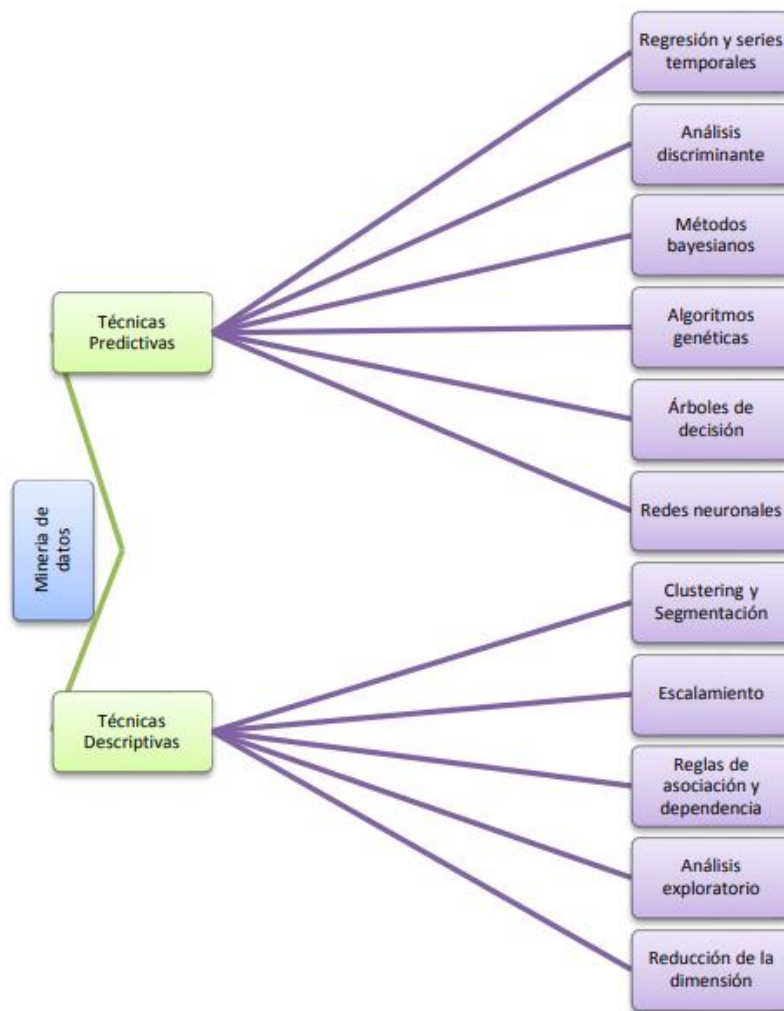


Figura 2.1 Técnicas y algoritmos en la Minería de Datos

2.9 CLUSTERIZACIÓN

Según (Wanner, 2004) , la clusterización o agrupamiento es una división de datos en grupos de objetos similares. Cada grupo (= un clúster) consiste en objetos que son similares entre sí mismos y diferentes a los objetos de otros grupos.

Desde la perspectiva del aprendizaje automático, la agrupación puede verse como un aprendizaje no supervisado de conceptos. Algunas de las aplicaciones son las siguientes:

- Minería de datos (análisis de ADN, estudios de marketing, estudios de seguros)
- Minería de texto (agrupación de tipos de texto)
- Recuperación de información (Agrupación de documentos)
- Lingüística computacional estadística (modelos de n-gramas basados en clústeres)
- Lexicografía Computacional Basada en Corpus

Existe una gran cantidad de algoritmos de clustering. En particular existen diferentes algoritmos basados en:

- Particiones
- Jerárquicos
- Densidades
- Rejillas
- Modelos
- Teoría de grafos
- Búsqueda combinatoria
- Técnicas Fuzzy
- Redes neuronales
- Kernels
- Para grandes conjuntos de datos (Morales & Escalante, 2017)

2.10 MATLAB

Es la abreviatura de *Matrix Laboratory* (laboratorio de matrices). Creado en 1984 por *MathWoks*. Combina un entorno de escritorio perfeccionado para el análisis iterativo y los procesos de diseño con un lenguaje de programación que expresa las matemáticas de matrices y *arrays* directamente.

Las apps de Matlab le permiten ver como funcionan diferentes algoritmos con sus datos. Realice iteraciones hasta obtener los resultados deseados y, después, genere automáticamente un programa de MATLAB para reproducir o automatizar su trabajo. Dentro de las aplicaciones de Matlab encontramos lo siguiente:

- Análítica de datos
- Comunicaciones Inalámbricas
- Aprendizaje profundo
- Visión artificial
- Procesamiento de señales
- Robótica
- Sistemas de control
- Finanzas Cuantitativas y Gestión de riesgos. (MathWoks, 2017)



Figura 2.2. Matlab (2018). MathWorks

2.11 PYTHON

Python es un lenguaje de programación interpretado, interactivo y orientado a objetos. Incorpora módulos, excepciones, tipado dinámico, tipos de datos dinámicos de muy alto nivel, y clases.

Este lenguaje combina potencia con una sintaxis muy clara. Tiene interfaces a muchas llamadas al sistema y bibliotecas, así como también a varios sistemas de ventanas, y es extensible en C o C++. Fue creado a principios de los 90 por Guido Van Rossum en el Stichting Mathematisch Centrum (Instituto Nacional de Investigación de Matemáticas y Ciencias de la Computación en Holanda), como sucesor de un lenguaje llamado ABC. En 2001 se fundó la Python Software Foundation (PSF), organización sin ánimo de lucro, para poseer la propiedad intelectual sobre Python. Posee las siguientes características:

- Extensible a otros sistemas software.
- Es orientado a objetos:
 - Herencia múltiple.
 - Ligadura dinámica.
 - Polimorfismo.
- Su núcleo es también orientado a objetos: jerarquía de clases.
- Portable: está implementado en C estándar usando E/S Posix.
- Gratuito, y de libre distribución. (Barranco, Guerrero Galindo, & Entrena, 2005)



Figura 2.3 Python (2018). Python.

2.12 SPYDER

Spyder es un potente entorno de desarrollo interactivo para el lenguaje Python con funciones avanzadas de edición, pruebas interactivas, depuración e introspección

en un entorno informático numérico gracias al soporte de IPython (intérprete interactivo mejorado de Python) y bibliotecas populares de Python como NumPy (álgebra lineal), SciPy (procesamiento de señal e imagen) o matplotlib (trazado interactivo 2D / 3D).

Spyder también se puede usar como una biblioteca que proporciona widgets potentes relacionados con la consola para sus aplicaciones basadas en PyQt; por ejemplo, se puede usar para integrar una consola de depuración directamente en el diseño de su interfaz gráfica de usuario. (Spyder, 2017)



Figura 2.4 spyder (2018). Spyder

2.13 WEKA

Software basado en un conjunto de librerías Java bajo licencia GPL, desarrollado en la Universidad de Waikato, de ahí el nombre de WEKA (*Waikato Environment for Knowledge Analysis*). Está orientado a la extracción de conocimiento a través de bases de datos con gran cantidad de información, contiene una gran colección de algoritmos y herramientas para analizar los datos junto con una interfaz sencilla, que hace que el usuario pueda usar este software de manera muy simple. (Navas Moreno, 2016)

Weka contiene las herramientas necesarias para realizar transformaciones sobre los datos, tareas de clasificación, regresión, clustering, asociación y visualización. Weka está diseñado como una herramienta orientada a la extensibilidad por lo que añadir nuevas funcionalidades es una tarea sencilla. (García Morate, 2017)



Figura 2.5 Weka (2018). Weka the university of waikato

2.14 INDICADORES DE RED

Para poder llegar a los objetivos de este trabajo, se requiere del uso de varios indicadores de red, algunos de ellos implícitos en la base de datos original, otros de ellos deberán ser programados para su obtención. Según (Bravo-Benitez, Kabadjova, & Martinez-Jaramillo, 2014) estos indicadores nos ayudarán a entender mejor el comportamiento del sistema.

2.14.1 Degree

El grado de un nodo (degree) en una red es una medida simple pero muy útil. Esta medida captura la cantidad de nodos a los que se conecta. El grado, $d(i)$, de un vértice, i , en un grafo se define como (Fórmula 2.1):

$$d_i = \sum_{j \in N(i)} a_{ij} \quad (2.1)$$

donde $N(i)$ es el conjunto de vecinos del vértice i ; es decir, el conjunto de vértices que tienen una conexión con el vértice i .

El grado interno (inner degree), d_i^- y el grado externo (outer degree), d_i^+ de un nodo, i , en un dígrafo se definen como (Fórmula 2.2):

$$d_i^- = \sum_{j \in N^-(i)} a_{ij} \quad \text{and} \quad d_i^+ = \sum_{j \in N^+(i)} a_{ij} \quad (2.2)$$

donde $N^-(i)$ es el conjunto de vecinos internos del vértice i , que es el conjunto de nodos que tienen un arco que termina en el nodo i . $N^+(i)$ es el conjunto de vecinos

externos de vértice i , el conjunto de nodos que tienen un arco que comienza en el nodo i .

2.14.2 Closeness

La centralidad de proximidad (Closeness centrality) tiene una interpretación de la independencia en las redes sociales en términos de control de comunicación. Un nodo con alto closeness dependería menos de otros nodos intermediarios para recibir mensajes. En el contexto del riesgo sistémico y el contagio financiero, esta medida puede ser asociado con la capacidad de un nodo para propagar el contagio, como tal un nodo está cerca del resto de la red. Se define como (Fórmula 2.3):

$$C_C(v) = \sum_{j \in V \setminus \{v\}} \frac{1}{d_G(v, j)}. \quad (2.3)$$

En el que $d_G(v, j)$ indica la longitud de la ruta más corta que une v y j .

2.14.3 Betweenness

La centralidad de intermediación (Betweenness), en las redes sociales, se asocia con ser estratégicamente ubicado en las rutas de comunicación de muchos otros nodos en la red. Un nodo con alta centralidad de intermediación tendría una importante influencia en otros nodos, ya que puede detener o distorsionar la información que pasa a través de él. Esta medida de centralidad es particularmente relevante en redes de sistemas de pago.

Estableciendo $\sigma_{ij} = \sigma_{ji}$ y denotando el número total de caminos más cortos entre i y j . Y que $\sigma_{ij}(v)$ sea el número total de caminos más cortos entre i y j que pasan a través del vértice v , entonces (Fórmula 2.4):

$$C_B(v) = \sum_{i \neq v \neq j \in V} \frac{\sigma_{ij}(v)}{\sigma_{ij}}. \quad (2.4)$$

2.14.4 Affinity

La afinidad (affinity) es una medida que, en función del grado de un nodo, describe el tipo de nodos a los que dicho nodo tiende a tener un enlace. (Fórmula 2.5)

$$a_i = \frac{1}{d_i} \sum_{j \in \mathcal{N}(i)} d_j \quad (2.5)$$

Si a_i está aumentando con d_i , entonces los nodos con alto grado tienden a tener relaciones con nodos que poseen un grado similar. Si a_i disminuye con d_i , entonces la mayoría de los vecinos de nodos de alto grado tienen menor grado.

Por el contrario, los nodos con bajo grado tienden a tener relaciones con nodos de alto grado.

Esta medida describe si los nodos en una red tienden a tener relaciones con nodos de grado similar o nodos con un grado diferente.

2.15 ALGORITMO DE CLUSTERIZACIÓN: DBSCAN

DBSCAN (*Density Based Spatial Clustering of Applications with Noise*) según (Pascual González, 2010) es uno de los primeros algoritmos de agrupamiento empleados en el enfoque de algoritmos basados en densidad. Su funcionamiento comienza seleccionando un punto t arbitrario, si t es un punto central, se empieza a construir un clúster alrededor de él, tratando de descubrir componentes denso-conectadas; si no, se visita otro objeto del conjunto de datos.

Identifica clústeres en grandes conjuntos de datos espaciales tomando en cuenta la densidad local de los elementos de un dataset, utilizando sólo un parámetro de entrada. DBSCAN también puede determinar qué información debe clasificarse como ruido o valores atípicos. A pesar de ello, su proceso de trabajo es rápido y se ajusta muy bien al tamaño de la base de datos, casi de forma lineal.

Un software que implementa el algoritmo DBSCAN es WEKA.

Algunas de las aplicaciones prácticas del algoritmo DBSCAN son:

- Imágenes de satélites.
- Cristalografía de rayos X.
- Detección de anomalías en datos de temperatura.
- Detección de anomalías en series temporales. (Backlund, Hedblom, & Neijman, 2011)

El algoritmo que se describe anteriormente tiene la capacidad de determinar automáticamente el número de grupos naturales, además de que identifica grupos irregulares y funciona para datos que algoritmos como el k-means no puede agrupar de manera satisfactoria, dicho algoritmo trabaja con base en dos parámetros lo cual lo hace un algoritmo simplificado y fácil de implementar, de acuerdo a la literatura dicho algoritmo es uno de los mejores en lo que se refiere a pruebas de tiempo.

CAPITULO 3

ESTADO DEL ARTE

3.1 CLASIFICACIÓN DE TÉCNICAS DE IDENTIFICACIÓN Y AGRUPACIÓN DE DATOS ANÓMALOS

La minería de datos no ayuda a extraer el conocimiento a partir de los datos que una empresa ha generado durante el tiempo de su existencia. Dentro de las aplicaciones dadas a la MD destacan las siguientes:

- Análisis de fidelización de clientes.
- Segmentación de Mercados.
- Optimización de la cadena de suministro.
- Detección y prevención de fraudes.
- Detección de intrusiones en sistemas computacionales.
- Detección de Anomalías.

La detección de anomalías, es conocida también como detección de desviaciones (*outliers*) según (Santamaría Ruíz, 2010), ya que define que objetos anómalos toman valores con una desviación significativa respecto a los valores típicos esperados, en diversas ocasiones los valores atípicos son frecuentemente tratados como ruido o error en las operaciones tales como las de clustering.

Para esto, se requiere identificar y agrupar los objetos o datos anómalos, clasificándose en técnicas basadas en:

Modelos: Se basa en el campo de la estadística, dando la pauta del conocimiento y distribución de los datos.

Proximidad: Se basa en la distancia de dichos objetos, entre mayor sea la distancia del objeto respecto a los demás se le considera como una anomalía.

Densidad: Estimación de densidad de los objetos, aquellos objetos localizados en regiones de baja densidad y distantes de sus vecinos, se considera anómalos. La principal característica de ésta técnica es que son de aprendizaje no supervisado, es decir, no se conoce la clase y se asigna una calificación a la instancia que refleja el grado con el cual la distancia es anómala.

3.2 CLUSTERING COMO HERRAMIENTA EN LA DETECCIÓN DE ANOMALÍAS

El clustering consiste en la agrupación de datos, es un tipo de aprendizaje no supervisado, debido a que no cuenta con clases preestablecidas del dataset de entrenamiento, por lo que se basa en la similitud de los atributos de los distintos datos. El agrupamiento de estos datos identifica clústers o regiones densamente pobladas tomando en cuenta medidas de distancia dentro de un dataset multidimensional. (M. Chen and J. Han, 1996).

Para el análisis de clustering existen los siguientes métodos:

- Jerárquicos
- Partición
- Basados en Densidad
- Basados en Cuadrículas
- Basados en Restricciones
- Escalables

Los algoritmos de clusterización funcionan con una metodología basada en la construcción inicial de un gran clúster y posteriormente la subdivisión del mismo hasta encontrar grupos de muestras más cercanas. (Santamaría Ruíz, 2010) nos dice que los diferentes tipos de clúster que se tienen son:

- Clúster bien separado. Esta definición idealista parte del hecho que todos los objetos de un grupo deben ser suficientemente similares.
- Clúster basado en el centro. Un clúster es un conjunto de objetos en el que un objeto está más cerca al centro del clúster, que al centro de otro clúster.
- Clúster contiguo. Un clúster es un conjunto de puntos, donde un punto en el clúster está más próximo a otro punto o puntos del clúster, que a cualquier otro punto que no pertenezca al clúster.

- Clúster basado en densidad. Este tipo de agrupamiento, se basa en el hecho de tener grupos en regiones de alta densidad, separados por regiones de baja densidad.
- Cluster de propiedad o Conceptual. Son clusters que tienen propiedad compartida o representan un concepto particular, es decir, hay puntos en común entre dos grupos. (Han & Kamber, 2006) (Ning Tan, Steinbach, & Kumar, 2006)

3.3 CLUSTERING BASADOS EN DENSIDAD

Los algoritmos de agrupamiento basados en densidad, obtienen un conjunto de datos de tal modo que los grupos se conforman por zonas de alta densidad, separados entre sí por áreas de baja densidad según (Benítez Sánchez, 2005) y (Pascual González, 2010) algunos de los algoritmos basados en densidad son:

DBSCAN: Se aplican a un espacio euclidiano 2D y 3D, en este algoritmo la idea principal es que para cada punto de un grupo, la vecindad debe contener al menos un número mínimo de puntos. (Ester, Krieger, Sander, & Xu, 1996)

OPTICS: Generalización del DBSCAN no escoge un valor apropiado para el parámetro del rango, y produce un resultado jerárquico relacionado con la conexión del agrupamiento. (Wikiwand, 2018)

KNN: Es un algoritmo de clasificación supervisada, calcula la distancia entre conjunto de entrenamiento y elementos del conjunto, selecciona los K ejemplos con menor distancia. (Uriz Martin & Galar Idoate, 2015)

DENCLUE (DENSity-based CLUstering): Define dos tipos de clústeres, es decir, centros definidos y clústeres definidos de varios centros. Es más rápido que DBSCAN. Su desventaja es que se debe seleccionar cuidadosamente su parámetro de densidad y umbral de ruido, pues puede afectar el resultado. (Shah, Napanda, & D'Mello, 2015)

3.5 DBSCAN APLICADO EN LA DETECCIÓN DE ANOMALÍAS

En el artículo Análisis de los métodos para la detección de anomalías en redes sociales según (Prado-Romero & Gago-Alonso, 2015) detectan grupos de forma arbitraria y solo agrupan una parte de los elementos D, el cual lo consideran como ruido. Las redes sociales son un dominio en el que es común encontrar conjuntos de datos de grandes dimensiones, es por eso que además de eficacia, los algoritmos aplicados a este estudio deben ser eficientes. Desde el punto de vista de la detección de anomalías los elementos considerados por el algoritmo DBSCAN serían considerados anómalos. El algoritmo recibe parámetros otorgados por el usuario, lo cual se toma como una desventaja, ya que como se determina de manera global y no en dependencia de densidad en las zonas donde se analizó el conjunto de datos, el algoritmo no es capaz de detectar anomalías puntuales locales, estas anomalías en redes sociales puede ser utilizada para la detección de fraude en redes telefónicas, las cuales son un caso particular de redes sociales. El descubrimiento de anomalías puntuales locales puede ser utilizado para encontrar aquellos usuarios del sector doméstico que reciben muchas llamadas telefónicas y no emiten ninguna, lo cual puede ser indicativo de ciertos tipos de fraude como el llamado bypass .

Otra aplicación de este algoritmo lo encontramos en *Anomaly detection in temperature data using DBSCAN algorithm* , este trabajo habla sobre en como la detección de anomalías ha impactado en los estudios climáticos, para detectar condiciones climáticas anormales causadas por el calentamiento global, describe como en una serie de tiempo de datos en la temperatura se observan movimientos altos y bajos. Determina los comportamientos normales de las temperaturas en verano e invierno. Los resultados obtenidos en DBSCAN son comparados con resultados de detección de anomalías estadísticos (Dadaser-Çelík, Çelík, & Sakir Dokuz, 2011)

Otro caso de investigación aplicando el algoritmo DBSCAN tenemos el de “Detección de patrones de distribución en ecología marina” en el cual intervienen un

gran conjunto de datos, los cuales sirven para modelar diferentes realidades y fenómenos existentes, en este trabajo se aborda la importancia de los avances tecnológicos y la expansión en la utilización de dispositivos tales como GPS o herramientas de tele-detección, los cuales generan información e incrementan su volumen debido a los diferentes fenómenos de la naturaleza geográfica los cuales van en aumento. Debido a esto los sistemas de información geográfica (GIS) se han convertido en más necesarios, para lo cual la utilización de los mismos han aumentado generando así el empleo de nuevos tipos de datos que dichos sistemas manejan, lo cual aumenta el espacio físico lo que conlleva a tomar los datos necesarios para modelar los diferentes fenómenos para así crear patrones de comportamiento en la información lo que apertura el campo de estudio de la minería de datos espacial, debido a esto surge la utilización del algoritmo DBSCAN que es utilizado para descubrir diferentes tipos de agrupaciones en datos espaciales, teniendo la ventaja de identificar el ruido entre los datos, obteniendo agrupaciones densas y relegando al instancias aisladas, además se considera que presenta un menor tiempo de ejecución con respecto al algoritmo CLARANS o K-means también probados en este trabajo. (Pol'la, Buccella, Cechich, & Doldan, 2016).

Por otra parte el artículo “Aplicación de técnicas de detección de anomalías a escenarios de ciudades inteligentes” (Romero & Alonso, 2017) nos muestra en el presente trabajo la aplicación de la minería de datos para detectar a anomalías en grandes volúmenes de datos pertenecientes a una red o diversas fuentes, para lo cual se propone el uso del algoritmo DBSCAN por su rapidez y por su utilidad en la detección temprana de riesgos de seguridad. Para el preprocesamiento y procesamiento en la detección de anomalías fue utilizado R dado la versatilidad en la minería de datos. Los pasos utilizados para la detección de anomalías fueron:

- 1.- Definen los parámetros a utilizar en DBSCAN (mínimo de puntos y ϵ)
- 2.- Limpieza de los datos eliminando valores nulos o datos incompletos
- 3.-Normalización de los datos entre 0 y 1
- 4.- Implementación del algoritmo con los parámetros establecidos.

CAPITULO 4

EXPERIMENTACIÓN

4.1 INTRODUCCIÓN

En el presente capítulo se muestran todas las etapas empleadas en el pre procesamiento del dataset y el procesamiento del mismo, aplicando una serie de técnicas que permitieran la eliminación de ruido en los datos, esto con la finalidad de reunir todos los elementos del conjunto de datos necesarios para el análisis y la experimentación llevada a cabo en esta sección, así mismo se realiza la limpieza de los datos que pueda ser de influencia negativa en los resultados que deseamos obtener, todo lo anterior fue elaborado utilizando Matlab y posteriormente migrado a Python para la aplicación del algoritmo pertinente.

La idea principal en lo que se presenta, es la obtención primeramente del comportamiento que tiene cada uno de los bancos actualmente, para posteriormente hacer un cálculo de distancias, realizar una comparación y así poder determinar mediante la aplicación del algoritmo si existieron o no datos anómalos al término del ejercicio.

4.2 DATASET ORIGINAL

Actualmente se cuenta para el análisis y manipulación un Dataset, el cuál fue proporcionado por el Banco de México, dicho conjunto de datos son emitidos por el SPEI en el periodo de tiempo comprendido de Agosto del 2004 a Diciembre del 2005, para su visualización y manejo estos archivos fueron trasladados a Matlab teniendo una extensión (.mat). En la Figura 4.1 se enlistan los archivos anteriormente mencionados:

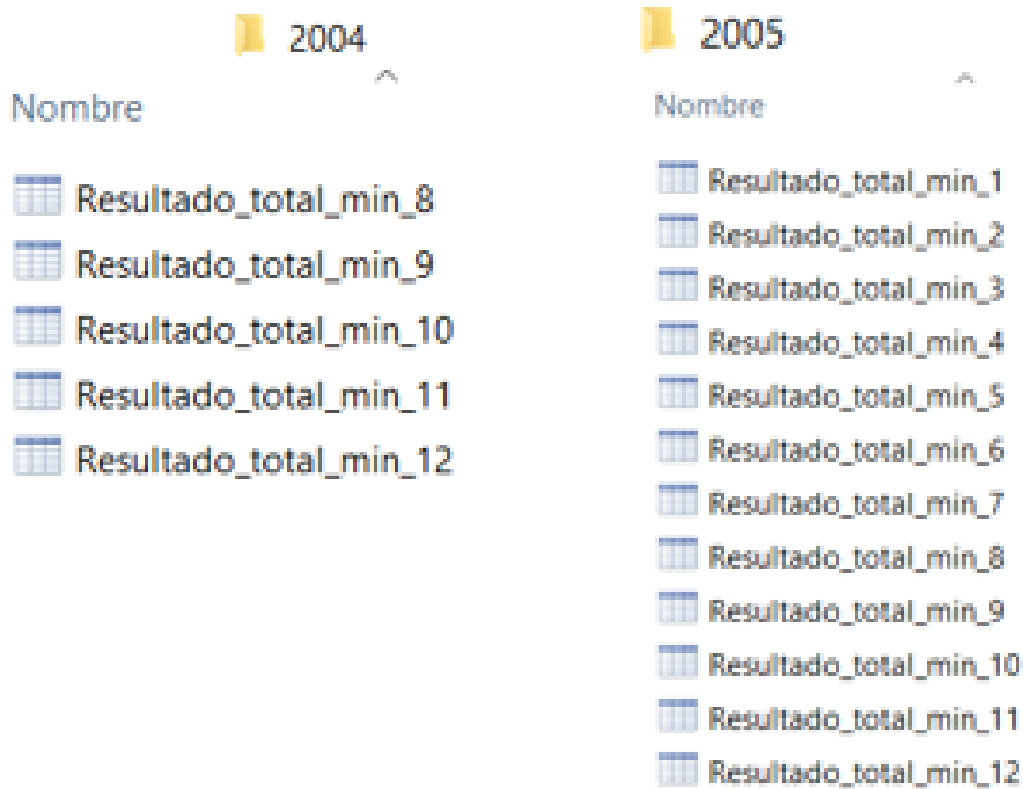


Figura 4.1. Archivos mensuales de Matlab Agosto 2004- Diciembre 2005

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

Cada elemento mostrado en la Figura 4.1 tiende a tener un tamaño de $1 \times N$, donde N es el número total de días que trabajó el SPEI en el mes que se esté analizando.

Cada uno de los archivos contiene 10 atributos (campos) siguientes:

- Saldos
- Mínimo
- Sumas
- Payments
- Payments_num
- Minimos_hora
- Suma_Hora
- S_Institution
- Reserve
- Nombre_i

Para la elaboración del trabajo presente se extrajeron por consiguiente los campos a utilizar los cuales son:

- Payments
- Payments_num
- Nombre_i

Payments contiene las matrices diarias originales de tamaño 130×130 de bancos contra bancos, dentro de los cuales comprende el monto total de cada envío en el día, por otra parte *payments_num*, es un sinónimo de *payments*, solo que ésta tiene almacenada la información de la cantidad de pagos, por su parte *nombre_j* contiene el nombre del día que se tiene almacenado del SPEI.

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

Vistos desde Matlab, los archivos diarios del SPEI de la estructura original se muestran como la siguiente figura 4.2:

Fields	saldos	minimo	sumas	payments	payments_num	minimos_hora	suma_hora	S_Institution	reserve	nombre_i
1	130x1 double	130x1 double	130x1 double	130x130 dou...	130x130 double	130x24 double	130x24 double	1x130 struct	1x1 struct	'13-08-2004....
2	130x1 double	130x1 double	130x1 double	130x130 dou...	130x130 double	130x24 double	130x24 double	1x130 struct	1x1 struct	'16-08-2004....
3	130x1 double	130x1 double	130x1 double	130x130 dou...	130x130 double	130x24 double	130x24 double	1x130 struct	1x1 struct	'17-08-2004....
4	130x1 double	130x1 double	130x1 double	130x130 dou...	130x130 double	130x24 double	130x24 double	1x130 struct	1x1 struct	'18-08-2004....
5	130x1 double	130x1 double	130x1 double	130x130 dou...	130x130 double	130x24 double	130x24 double	1x130 struct	1x1 struct	'19-08-2004....
6	130x1 double	130x1 double	130x1 double	130x130 dou...	130x130 double	130x24 double	130x24 double	1x130 struct	1x1 struct	'20-08-2004....
7	130x1 double	130x1 double	130x1 double	130x130 dou...	130x130 double	130x24 double	130x24 double	1x130 struct	1x1 struct	'23-08-2004....
8	130x1 double	130x1 double	130x1 double	130x130 dou...	130x130 double	130x24 double	130x24 double	1x130 struct	1x1 struct	'24-08-2004....
9	130x1 double	130x1 double	130x1 double	130x130 dou...	130x130 double	130x24 double	130x24 double	1x130 struct	1x1 struct	'25-08-2004....
10	130x1 double	130x1 double	130x1 double	130x130 dou...	130x130 double	130x24 double	130x24 double	1x130 struct	1x1 struct	'26-08-2004....

Figura 4.2 Estructura interna de los archivos diarios SPEI

4.3 PRE PROCESAMIENTO DEL DATASET

4.3.1 Unión de base de datos diaria y eliminación de campos innecesarios

Como primera etapa se considera unificar en una estructura elaborada en Matlab todos los archivos obtenidos anteriormente, permaneciendo solo con las columnas necesarias para el trabajo como lo son: payments, payments_num y nombre_j,, una vez realizado esto, se obtiene la base de datos final con la cual se va a trabajar. Un fragmento del código realizado en Matlab utilizado se muestra en la Figura 4.3

```
entrada='C:\Escritorio\Abigail\SPEI\datosin';
salida='C:\Escritorio\Abigail\SPEI\datosout';
cd(entrada);
Dir1_meses=dir;
numMeses=length(Dir1_meses);
num_inst=98;
k=1;
] for mes=3:numMeses
    nombrei=Dir1_meses(mes).name;
    load(nombrei);
    numd=size(Resultado_total_min,2);
    aux=100;
] for x=1:numd
    nombre=strcat('dia',Resultado_total_min(x).nombre_i);
    numpagos= Resultado_total_min(x).payments_num;
    pagos= Resultado_total_min(x).payments;
    sum=Resultado_total_min(x).sumas;
    hora=Resultado_total_min(x).suma_hora;
    estructuraNueva(k)=struct('payments',pagos,'payments_num',numpagos,'por_hora',hora,'di
    k=k+1;
    aux=aux+1;
- end
clear numd;
-end
cd(salida);
save(strcat('estructura_agosto2004_dic_2005_reducida'),'estructuraNueva');
```

Figura 4.3 Código utilizado para la unión de base de datos y eliminación de campos innecesarios

4.3.2 REDUCCIÓN DE BANCOS

Como se mencionó anteriormente los archivos son emitidos por el SPEI para el tratamiento de sus datos, una de los puntos importantes es una matriz de 130 con la que cuenta la estructura originalmente proporcionada, que hace referencia a los 130 bancos participantes, sin embargo, no todos estos participantes laboraban en el SPEI en el periodo comprendido, es por eso que muchos de los valores expuestos en la matriz están en ceros.

Es por eso que la segunda etapa consistió en una de las partes medulares del pre procesamiento, en la cual se removieron dichos valores con cero, extrayendo solo de las filas y columnas de los rangos 1 al 32 y 56 al 61 que son las que muestran la información a utilizar. En la siguiente figura se muestra el código utilizado para la eliminación del ruido que da lugar a nuestra reducción de bancos, los cuáles se comprimieron a solo 38 participantes.

```
load('estructura_agosto2004_dic_2005_reducida.mat');
datos=estructura2;
a=1:32;
b=56:61;
c=[a,b];
for i=1:355
    pag=datos(i).payments(c,c);
    pag_n=datos(i).payments_num(c,c);
    porhora=datos(i).por_hora(c,1:24);
    dia=datos(i).dia;
    datos_res(i)=struct('pagos',pag,'numnpagos',pag_n,'suma',porhora,'dias',dia);
end
save(strcat('nueva_estructura_38'),'datos_res');
```

Figura 4.4 Código utilizado para la reducción de bancos participantes durante el SPEI

4.3.3 OBTENCIÓN DE ARCHIVOS POR BANCO

A continuación del dataset se desean obtener archivos que estén seccionados por bancos, lo cual servirá como primera etapa para conocer su comportamiento posteriormente, como se muestra en el código a continuación se tienen los distintos archivos por bancos obtenidos:

```
load('nueva_estructura_38.mat');
datos=datos_res;

]for j=1:38
]   for i=1:355
       porhora=datos(i).porhoras(j,1:24);
       dia=datos(i).dias;
       nombre=strcat('banco',num2str(j));
       solo_hora(i)=struct(nombre,porhora,'dias',dia);
-   end
       save(strcat('todas_horas',num2str(j)), 'solo_hora');
       clearvars solo_hora
- end
```

Figura 4.5 Código utilizado para obtener los archivos de los bancos participantes.

La ejecución del código mostrado en la Figura 4.5 nos genera 38 archivos cada uno correspondiente a un banco como se ve en la siguiente Figura 4.6:

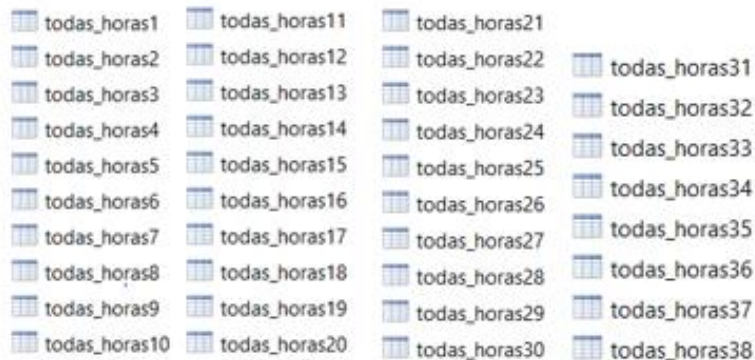


Figura 4.6 Archivos de los 38 bancos obtenidos

4.3.4 GENERAR INDICADORES

Se trabajó con una estructura final, dentro de la cual solo se tienen los bancos que nos interesan, así como los campos a utilizar los cuales son de payments y el nombre del día, el siguiente procedimiento es programar los indicadores de red necesarios. Para poder realizar dicho paso, se deberá crear primero una matriz de adyacencias, la cual contendrá un 1 para aquellos bancos quienes el banco emisor hace un envío a otro, y un 0 a aquellos bancos con los que tiene comunicación. Dicha matriz será denominada W_b , su aspecto será como la figura 4.7.

Emisores	0	1	1	0	0	1
	1	0	1	0	1	1
	0	1	0	1	0	1
	0	1	1	0	1	0
	1	1	1	0	0	1
	0	1	1	0	0	0
	Receptores					

Figura 4.7 Matriz de adyacencia

Ahora se procede a calcular la matriz de adyacencia W_l , que es la matriz de recepción de pagos. $W_l = W_b^T$. Figura 4.8

Receptores	0	1	0	0	1	0
	1	0	1	1	1	1
	1	1	0	1	1	1
	0	0	1	0	0	0
	0	1	0	1	0	0
	1	1	1	0	1	0
		Emisores				

Figura 4.8 Matriz de recepción de pagos

Finalmente, con las matrices W_l y W_b procederemos a calcular la matriz W quien será la matriz de pesos que indique el número de relaciones que tiene una entidad en ese día. Se expresa por $W = W_l + W_b$,

Participante	0	1	1	0	1	1
	1	0	1	1	1	1
	1	1	0	1	1	1
	0	1	1	0	1	0
	1	1	1	1	0	1
	1	1	1	0	1	0
	Participante comunicado					

Figura 4.9 Matriz W , número de relaciones de las entidades por día

Una vez calculados W , W_l y W_b se procede a calcular los degree. Como ya me mencionó anteriormente, el degree es una medida usada para calcular el número de nodos con que un elemento está conectado. El degree se encuentra definido cómo (Fórmula 4.9):

$$d_i = \sum_{j \in N(i)} a_{ij} \quad (4.9)$$

Donde $N(i)$ es el conjunto de vértices vecinos i (vértices que tienen una arista hacia i). Relacionado con la Figura 4.6 a continuación nos muestra el degree del ejemplo.

Degree	4
	5
	5
	3
	5
	4

Figura 4.10 Degree del ejemplo

Con el uso del degree, y con base en el conocimiento obtenido en el capítulo 2, se programaron los demás indicadores necesarios para la siguiente etapa, como lo son inner degree, outer degree, montos in, montos out, betweenness, affinity y

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

closeness, y se almacenaron en una nueva estructura de Matlab como se muestra en la siguiente figura.

```
datos=D_Resultado_matrices_2013_01_95;
for i=1:355
    nombre=extractBefore(extractAfter(datos(i).G.dias,1),11);
    degree=datos(i).D;
    degree_in=datos(i).D_in;
    degree_out=datos(i).D_out;
    montos_in=datos(i).Sl;
    montos_out=datos(i).Sb;
    affinity=datos(i).Aff;
    betweenness=datos(i).Betweenness;
    closeness=datos(i).Closeness;
    salida(i)=struct('nombre',nombre,'degree',degree,'degree_in',degree_in,'degree_out',...
        degree_out,'montos_in',montos_in,'montos_out',montos_out,'affinity',affinity,'betweenness',betweenness,'closeness',closeness);
end
save(strcat('testbdfinalabril'),'salida');
```

Figura 4.11 Código para la programación de indicadores.

Por último, se deberá dar el formato correspondiente de filas y columnas esperado como parámetro de entrada en Python, por lo tanto, se utilizará el siguiente código el cual generará un archivo .mat por cada día que laboró el SPEI.

```
for i=1:355
    fecha=datos(i).nombre;
    degree=datos(i).degree;
    degree_in=datos(i).degree_in;
    degree_out=datos(i).degree_out;
    montos_in=datos(i).montos_in;
    montos_out=datos(i).montos_out;
    affinity=datos(i).affinity;
    betweenness=datos(i).betweenness;
    closeness=datos(i).closeness;

    salidasIn(:,1)=degree';
    salidasIn(:,2)=degree_in';
    salidasIn(:,3)=degree_out';
    salidasIn(:,4)=montos_in';
    salidasIn(:,5)=montos_out';
    salidasIn(:,6)=affinity';
    salidasIn(:,7)=betweenness';
    salidasIn(:,8)=closeness';

    save(strcat('bdd_',int2str(i),'_',fecha),'salidasIn');
end
```

Figura 4.12 Código utilizado para el formato de indicadores

4.4 PROCESAMIENTO DEL DATASET

4.4.1 Procesamiento con Python

Es esta etapa se decidió utilizar el lenguaje de programación Python para un mejor análisis estadístico y sobretodo visual de los datos que fueron procesados previamente en Matlab, dichos datos fueron obtenidos y exportados de dicho banco de datos al formato csv, en el cual se realiza nuevamente una separación de archivos por banco. Dicho proceso se muestra a continuación, tomando como ejemplo el banco1:



Figura 4.13 Proceso de conversión de Matlab - Csv - Python

Primera etapa: Se tienen los datos originales obtenidas de la base de datos original, estos datos ya se encuentran pre procesados, todo esto se realizó en Matlab, A continuación se muestra un archivo del banco número 1 en el que sus filas son el número de días y las columnas son los indicadores utilizados para su procesamiento, los cuales son: degree, degreeIn, degreeOut, monOut, monIn, Affinity, betweness, día, mes y año.

salidas											
355x11 double											
	1	2	3	4	5	6	7	8	9	10	11
1	6	5	2	6.4644e+08	1.3845e+09		3.5567	14.0000	13	8	2004
2	11	8	7	2.4597e+09	4.0950e+09	8.8182	24.1250	16.3333	16	8	2004
3	10	10	5	2.3972e+09	5.1969e+09	10.6000	5.2436	17.3333	17	8	2004
4	10	6	8	1.9326e+09	3.8335e+09	10.1000	9.7346	17.5000	18	8	2004
5	11	10	4	2.3460e+09	5.7363e+09	12.8182	1.4409	18.8333	19	8	2004
6	8	8	4	5.0322e+09	5.9672e+09	10.8750	2.1512	16.8333	20	8	2004
7	7	7	2	1.3396e+09	2.8347e+09	10	11.8714	16.3333	23	8	2004
8	9	9	3	1.6491e+09	3.5123e+09	12	2.4038	18.1667	24	8	2004
9	12	7	10	1.8727e+09	4.3518e+09	9.2500	114.5539	20.1667	25	8	2004
10	13	10	7	2.1451e+09	4.6390e+09	10.6923	28.9603	19.8333	26	8	2004
11	9	8	4	2.3021e+09	5.8427e+09	12	4.6972	17.8333	27	8	2004
12	9	9	2	2.9969e+09	4.5323e+09	10.7778	26.0833	18.8333	30	8	2004
13	16	9	11	4.6732e+09	6.4087e+09	10	155.3410	23.5000	31	8	2004
14	10	10	6	2.6362e+09	5.2438e+09	10	89.0568	18.8333	1	9	2004
15	9	5	8	3.8958e+09	4.7531e+09	11	61.2528	19.5000	2	9	2004
16	10	9	6	6.1080e+08	3.9102e+09	12.2000	68.9347	19.8333	3	9	2004
17	12	8	9	2.4859e+09	4.8109e+08	7.7500	77.2881	18.1667	6	9	2004
18	20	11	13	4.2480e+09	8.9244e+09	11.1000	99.7688	25.3333	7	9	2004
19	23	13	18	3.1987e+09	5.0082e+09	10.6522	171.3179	27	8	9	2004
20	11	8	8	1.8231e+09	4.6500e+09	15.1818	0.8512	20.1667	9	9	2004
21	12	11	7	8.8855e+08	2.3101e+09	11.1667	13.5167	20.5000	10	9	2004
22	15	15	7	2.6413e+09	4.1618e+09	11.6000	23.8850	22	13	9	2004
23	12	10	6	3.8194e+09	6.1674e+09	13	7.7786	21.3333	14	9	2004
24	10	8	6	3.4306e+09	3.7240e+09	15.1000	1.6095	20.6667	15	9	2004

Figura 4.14 Estructura interna archivo Matlab, Banco 1

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

Segunda etapa: Los datos originales y pre procesados en Matlab, son convertidos a csv para posteriormente ser exportados a Python, el archivo .csv se muestra a continuación, aún en esta etapa los valores se mantienen intactos.

Degree	DegreeIn	DegreeOut	MonIn	MonOut	Affinity	Closeness	Betweenness	Día	Mes	Año
5	5	0	0	2.21E+09	9.2	1.5	13.333	13	8	2004
7	6	4	3.09E+08	8.58E+08	10	4.95	13.917	16	8	2004
11	7	8	7.06E+09	2.50E+09	9.9091	30.174	17.833	17	8	2004
12	5	10	3.74E+09	1.28E+09	9.8333	25.165	18.5	18	8	2004
17	11	13	1.33E+10	8.80E+09	9.8824	88.95	22	19	8	2004
12	12	6	6.99E+09	1.10E+10	9.8333	9.3833	19.167	20	8	2004
9	5	6	8.47E+09	3.78E+09	8.5556	62.029	17.667	23	8	2004
11	8	6	1.56E+09	6.21E+09	10	58.797	19.333	24	8	2004
13	9	12	1.07E+10	8.84E+09	11.462	0.975	20.667	25	8	2004
10	7	9	1.17E+10	8.94E+09	11.7	13.625	18.167	26	8	2004
9	9	1	3.60E+07	6.81E+09	12.667	1.4262	17.833	27	8	2004
11	11	2	6.75E+08	1.13E+10	11.182	5.5619	19.667	30	8	2004
14	12	7	8.85E+09	1.39E+10	12.571	27.107	22.5	31	8	2004
14	13	9	5.69E+09	9.87E+09	9.7857	25.618	20.833	1	9	2004
16	9	11	1.53E+10	6.38E+09	9.6875	60.663	23.5	2	9	2004
8	8	4	3.01E+08	3.85E+09	12.75	10.957	18.833	3	9	2004
10	9	5	5.18E+09	7.44E+09	9.5	13.95	17.333	6	9	2004
18	13	14	1.30E+10	2.22E+10	12.778	20.731	24.5	7	9	2004
16	14	10	7.30E+09	1.57E+10	14.125	13.965	23.5	8	9	2004
16	12	11	1.45E+10	2.11E+10	12.438	17.721	22.667	9	9	2004
16	13	8	2.19E+10	1.96E+10	11.813	24.151	22.5	10	9	2004
17	13	10	1.71E+10	2.84E+10	11.647	23.64	23	13	9	2004

Figura 4.15 Datos originales CSV- Banco1

Tercera etapa: Una vez convertidos los datos a .csv, se realiza la programación necesaria en Python para poder realizar el escalado de los datos resultantes el cual consiste en una estandarización de los resultados entre 0 y 1, esto debido a que los montos expresados son de cantidades elevadas y existe una diferencia considerable la cual bifurca mucha los resultados, también fueron utilizados los indicadores mencionados anteriormente, en este proceso si fue necesario considerar la columna correspondiente al mes, ya que el siguiente procedimiento se toma dicha columna para realizar la agrupación mensual necesaria. Se muestra a continuación el archivo de la tabla de los datos escalados arrojados por el algoritmo ejecutado en Python.

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

	0	1	2	3	4	5	6	7
0	0	0	0	0	0.00599392	0.0917815	0.00798297	0
1	0.0740741	0.037037	0.125	0.00138221	0	0.188956	0.0263438	0.030469
2	0.222222	0.0740741	0.25	0.0316301	0.00726959	0.177915	0.160585	0.234779
3	0.259259	0	0.3125	0.0167635	0.00186208	0.168707	0.133928	0.269578
4	0.444444	0.222222	0.40625	0.0596347	0.0352434	0.174671	0.47339	0.452183
5	0.259259	0.259259	0.1875	0.0312831	0.0448577	0.168707	0.0499377	0.304377
6	0.148148	0	0.1875	0.0378985	0.0129478	0.0135073	0.330117	0.226118
7	0.222222	0.111111	0.1875	0.00699722	0.023734	0.188956	0.312916	0.313038
8	0.296296	0.148148	0.375	0.047954	0.0354262	0.366543	0.00518893	0.382637
9	0.185185	0.0740741	0.28125	0.0525206	0.0358637	0.395452	0.072512	0.252204
10	0.148148	0.148148	0.03125	0.000160969	0.0264029	0.512912	0.00759021	0.234779
11	0.222222	0.222222	0.0625	0.00302055	0.0465305	0.332532	0.0296003	0.330464
12	0.333333	0.259259	0.21875	0.0396092	0.0577564	0.501251	0.144263	0.47827
13	0.333333	0.296296	0.28125	0.0254808	0.0400017	0.162925	0.136338	0.391298
14	0.407407	0.148148	0.34375	0.0684187	0.0244962	0.150997	0.322847	0.530443
15	0.111111	0.111111	0.125	0.0013489	0.0132908	0.522994	0.0583129	0.286952
16	0.185185	0.148148	0.15625	0.0231796	0.0291952	0.128222	0.0742416	0.208692
17	0.481481	0.296296	0.4375	0.0582333	0.0945224	0.526395	0.11033	0.582616
18	0.407407	0.333333	0.3125	0.0326625	0.0657965	0.690013	0.0743214	0.530443

Figura 4.16 Datos escalados primera parte, Banco 1

4.4.2 Ejecución de DBSCAN por banco

El objetivo fundamental en el que se basa la tesis presentada es la detección de anomalías en datos financieros, los cuáles están implícitos en la información que se mostró anteriormente tomando como ejemplo el banco 1. Para poder realizar dicho análisis y así conocer el funcionamiento que normalmente tiene un banco y detectar alguna posible anomalía, fue necesario agrupar mensualmente dichos datos obtenidos de los .csv y posteriormente escalados. Se muestra a continuación el código realizado en Python en el cual se obtienen los .csv, se procede a realizar el escalado y se ejecuta el algoritmo DBSCAN, dicho algoritmo como se estudia en la teoría requiere 2 parámetros, los cuales son: **eps** (que se define como la distancia que existe del radio de un punto hacia otro punto más cercano) y el **minPts** (el cual es el número de objetos que conforman un cluster), para las pruebas realizadas en la fase de experimentación se tomaron los siguientes valores:

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

Tabla 4.1

Parámetros utilizados en el algoritmo DBSCAN para la fase de experimentación

Eps	MinPts
0.3	5

Dichos valores se tomaron a partir de diversos resultados obtenidos en la literatura, en la cual expresa los parámetros utilizados adecuadamente y los tipos de clúster que podemos obtener como resultado, en la siguiente tabla se expresan dichos resultados.

Tabla 4.2

Evaluación de Resultados de acuerdo a sus parámetros

Eps	MinPts	Resultado
Alto	Alto	Pocos clusters, grandes y densos
Bajo	Bajo	Muchos clusters, pequeños y menos densos
Alto	Bajo	Clusters Grandes y menos densos
Bajo	Alto	Clusters Pequeños y densos

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

```
10 import pandas as pd
11 from sklearn.preprocessing import MinMaxScaler
12 from sklearn import metrics
13
14 attributes = ['A1', 'A2', 'A3', 'A4', 'A5', 'A6', 'A7', 'A8', 'A9', 'A10', 'A11']
15 banco="1"
16 archivo="C:/Users/Abii Echeverria/Desktop/Datos_banco/datos_originales/bancocsv_"+banco+".csv"
17 data = pd.read_csv(archivo, sep=',', header=None, names=attributes)
18
19 data2=data.drop("A9",1)
20 data2=data2.drop("A10",1)
21 data2=data2.drop("A11",1)
22
23
24 escalador = MinMaxScaler()
25 df_escalado = escalador.fit_transform(data2)
26
27 from sklearn.cluster import DBSCAN
28
29 db = DBSCAN(eps=.3, min_samples=5).fit(df_escalado)
30 core_samples_mask = np.zeros_like(db.labels_, dtype=bool)
31 core_samples_mask[db.core_sample_indices_] = True
32 labels = db.labels_
33
34 n_clusters_ = len(set(labels)) - (1 if -1 in labels else 0)
35
36
37 import matplotlib.pyplot as plt
38
39 unique_labels = set(labels)
40 colors = [plt.cm.Spectral(each)
41           for each in np.linspace(0, 1, len(unique_labels))]
42
43 for k, col in zip(unique_labels, colors):
44     if k == -1:
45         # Black used for noise.
46         col = [0, 0, 0, 1]
47
48     class_member_mask = (labels == k)
49
50     xy = df_escalado[class_member_mask & core_samples_mask]
51     plt.plot(xy[:, 0], xy[:, 1], 'o', markerfacecolor=tuple(col),
52             markeredgecolor='k', markersize=14)
53
54     xy = df_escalado[class_member_mask & ~core_samples_mask]
55     plt.plot(xy[:, 0], xy[:, 1], 'o', markerfacecolor=tuple(col),
56             markeredgecolor='k', markersize=6)
57
58 plt.title('Estimated number of clusters: %d' % n_clusters_)
59 plt.savefig("banco"+banco+".png")
60 plt.show()
```

Figura 4.17 Algoritmo Python con escalado de datos y ejecución del algoritmo DBSCAN

4.5 ANÁLISIS DE BANCOS

Al ejecutar el algoritmo DBSCAN nos da como resultado los días anómalos teniendo como referencia los 355 días que trabajó el banco, así también nos muestra el clúster del cual forma parte en cada uno de sus días teniendo como número de identificación el 0, para el clúster '0', el 1 para el clúster '1' y el -1 para mostrar las anomalías. En la siguiente figura 4.18 se presentan algunos días de los 355 días con su diferente clasificación tomando en cuenta que su contador inicia en el 0 que corresponde al día 1, dichos resultados fueron obtenidos de Python.

	0		0		0		0		0		0		
0	0	19	0	38	0	57	0	76	0	95	0	114	0
1	0	20	0	39	0	58	0	77	0	96	0	115	0
2	0	21	0	40	0	59	0	78	0	97	0	116	0
3	0	22	0	41	0	60	0	79	0	98	0	117	0
4	0	23	0	42	0	61	0	80	0	99	0	118	0
5	0	24	0	43	0	62	0	81	0	100	0	119	0
6	0	25	0	44	0	63	0	82	0	101	0	120	0
7	0	26	0	45	0	64	0	83	0	102	0	121	-1
8	0	27	0	46	0	65	0	84	0	103	0	122	0
9	0	28	0	47	0	66	0	85	0	104	0	123	0
10	0	29	0	48	0	67	0	86	0	105	0	124	0
11	0	30	0	49	0	68	0	87	0	106	0	125	0
12	0	31	0	50	0	69	0	88	0	107	0	126	0
13	-1	32	0	51	0	70	0	89	0	108	0	127	-1
14	-1	33	0	52	0	71	0	90	0	109	0	128	0
15	0	34	0	53	0	72	0	91	0	110	0	129	0
16	0	35	0	54	0	73	-1	92	0	111	0	130	0
17	0	36	0	55	0	74	0	93	0	112	0	131	0
18	0	37	0	56	0	75	0	94	0	113	0	132	0

Figura 4.18 Análisis banco 10. Clasificación de días con sus clúster y sus respectivas anomalías.

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

Tomando como referencia los resultados obtenidos de los días en la figura anterior, a continuación en esta otra figura, se muestra un ploteo de las anomalías surgidas en los días que se mencionaron. Teniendo en el eje de las X los días analizados y en el eje de las Y el número de clúster al que pertenece cada uno de los días del banco 10. Las demás gráficas se muestran en el anexo 1.

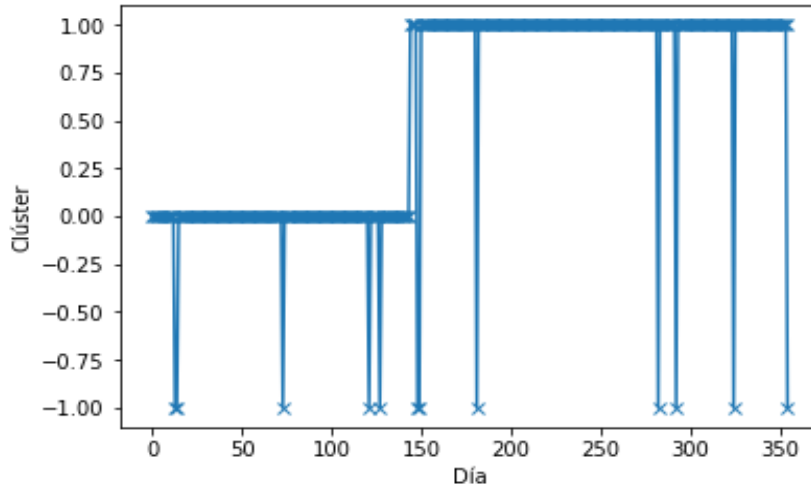


Figura 4.19. Gráfica de resultados anomalías obtenidas. Banco 1

Una vez obtenidos los días anómalos (*outliers*) de la Figura anterior, se presenta a continuación un análisis de los resultados, todo con respecto al banco número 10. En dicho análisis se enlistan los indicadores de red sobre los cuáles se trabajaron los cuales son: degree, degreeIn, degreeOut, montosIn, montosOut, Betewwness, Closeness, y affinity.

BANCO 10							
Días							
	01-sep-04	02-sep-04	25-nov-04	01-feb-05	09-feb-05	10-mar-05	11-mar-05
Anomalías	14	15	74	122	128	149	150
Degree	5	3	4	9	8	17	16
DegreeIn	3	3	0	4	3	13	7
DegreeOut	2	0	4	7	7	13	15
MontosIn	*****						
MontosOut	*****						
Betweenness	8.4	11	7	18.444	16.125	20.294	20.75
Closeness	20.542	11.583	8.8667	6.5354	5.9269	5.5222	4.4591
Affinity	15.833	15.167	16.833	20.333	19.833	24.5	24

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

	29-abr-05	20-sep-05	04-oct-05	17-nov-05	30-dic-05
Anomalías	182	283	293	325	355
Degree	13	10	10	16	7
DegreeIn	9	9	8	7	7
DegreeOut	12	9	7	16	6
MontosIn	*****				
MontosOut	*****				
Betweenness	23	25.3	24.3	21.063	28
Closeness	1.6385	0.1	0.1	1.3937	0
Affinity	22.5	21	21	24	19.5

Tabla 4.3 Análisis banco 10. Identificación de anomalías por indicador dependiendo su media y desviación estándar

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

Los valores que se muestran a continuación corresponden a las medias y desviaciones estándar de cada uno de los clusters obtenidos de ese banco, los cuales nos ayudaron a comprobar el motivo por el cual el algoritmo DBSCAN identificó estos días como anómalos. Como se puede observar en este mes para que un día sea considerado anómalo el algoritmo identificó efectivamente al menos 2 de sus indicadores los cuáles quedan fuera del rango en ambos clúster.

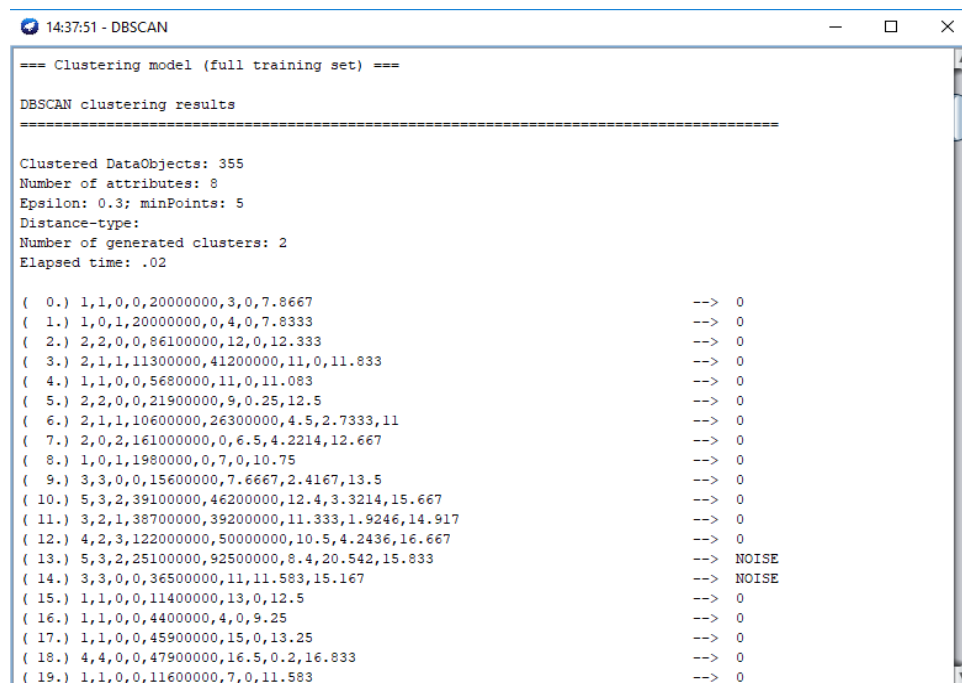
	Media C0	DE C0	Media C1	DEC1
Degree	3.172661871	1.631207	11.480392	1.601044
DegreeIn	6.371830986	1.368420	3.565452	1.297011
DegreeOut	6.154929577	1.141507	4.361961	1.627746
MontosIn	*****			
MontosOut	*****			
Betweenness	20.86355972	4.837401	5.042866	1.408886
Closeness	0.869482434	1.765650	1.838550	0.682367
Affinity	19.46251268	2.286255	3.193482	0.826566

Tabla 4.4 Media y desviación estándar obtenida por cada cluster. Comprobación de días anómalos.

4.6 COMPROBACIÓN DE RESULTADOS CON WEKA

Una vez finalizado los resultados obtenidos del algoritmo DBSCAN en Python, se corroboró estadísticamente el motivo por el cuál algunos días fueron anómalos, cabe recalcar que no precisamente dichos días anormales contienen a ser algo malo, simplemente surgieron como motivo de un evento fuera de lo normal en su curso con respecto a los demás días.

El banco número 10, así como otros que se presentan más adelante sirvieron para confirmar dichos resultados. Para la comprobación del algoritmo implementado se decidió utilizar un software y realizar la ejecución de dicho algoritmo con el mismo banco en el cual se comprobaron los días anómalos. El resultado fue favorable y sirve como comprobación y respaldo de los resultados obtenidos. A continuación se muestra un ejemplo en la Figura 4.20 que la fila con el número 13, 14, lo cual corresponde en días al 14 y 15 muestran un dato con ruido “noise” lo cual corresponde a una anomalía, dichos días también fueron determinados con anterioridad en la Figura 4.18 donde dicho dato afectado tiene una etiqueta de -1, que presenta claramente un día fuera de lo habitual.



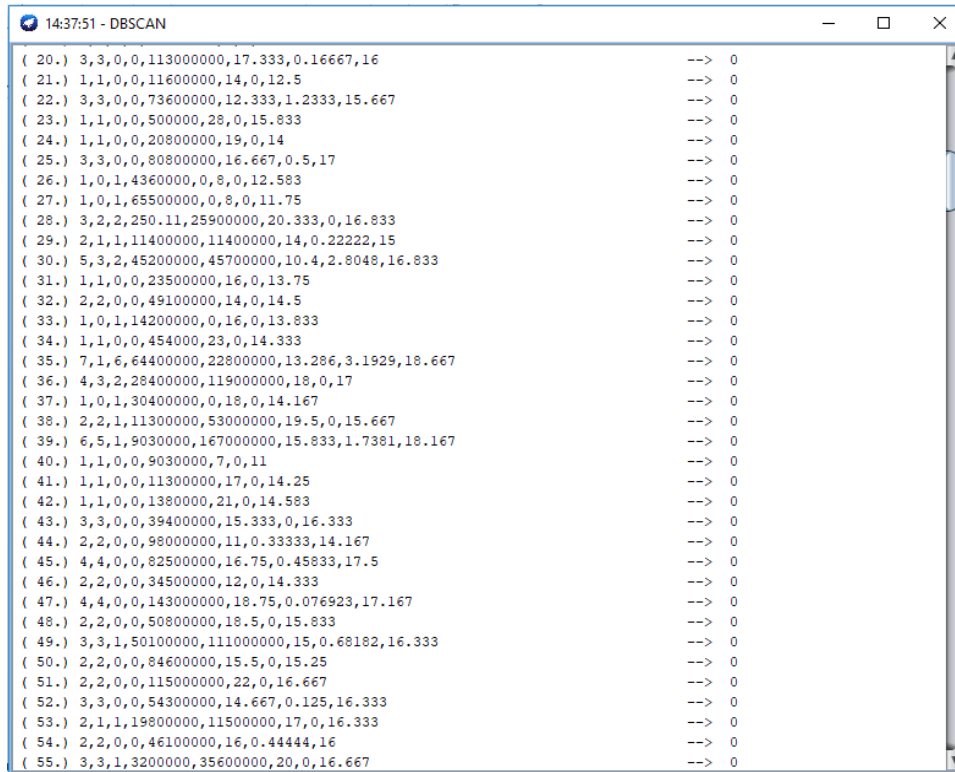
```
14:37:51 - DBSCAN
=== Clustering model (full training set) ===

DBSCAN clustering results
-----

Clustered DataObjects: 355
Number of attributes: 8
Epsilon: 0.3; minPoints: 5
Distance-type:
Number of generated clusters: 2
Elapsed time: .02

( 0.) 1,1,0,0,20000000,3,0,7.8667      --> 0
( 1.) 1,0,1,20000000,0,4,0,7.8333      --> 0
( 2.) 2,2,0,0,86100000,12,0,12.333    --> 0
( 3.) 2,1,1,11300000,41200000,11,0,11.833 --> 0
( 4.) 1,1,0,0,5680000,11,0,11.083     --> 0
( 5.) 2,2,0,0,21900000,9,0,25,12.5    --> 0
( 6.) 2,1,1,10600000,26300000,4,5,2.7333,11 --> 0
( 7.) 2,0,2,161000000,0,6,5,4,2214,12.667 --> 0
( 8.) 1,0,1,1980000,0,7,0,10.75      --> 0
( 9.) 3,3,0,0,15600000,7,6667,2.4167,13.5 --> 0
(10.) 5,3,2,39100000,46200000,12,4,3.3214,15.667 --> 0
(11.) 3,2,1,38700000,39200000,11,333,1.9246,14.917 --> 0
(12.) 4,2,3,122000000,50000000,10,5,4,2436,16.667 --> 0
(13.) 5,3,2,25100000,92500000,8,4,20.542,15.833 --> NOISE
(14.) 3,3,0,0,36500000,11,11.583,15.167 --> NOISE
(15.) 1,1,0,0,11400000,13,0,12.5     --> 0
(16.) 1,1,0,0,4400000,4,0,9.25      --> 0
(17.) 1,1,0,0,45900000,15,0,13.25    --> 0
(18.) 4,4,0,0,47900000,16,5,0,2,16.833 --> 0
(19.) 1,1,0,0,11600000,7,0,11.583    --> 0
```


IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN



```
14:37:51 - DBSCAN
( 20.) 3,3,0,0,113000000,17.333,0.16667,16 --> 0
( 21.) 1,1,0,0,116000000,14,0,12.5 --> 0
( 22.) 3,3,0,0,736000000,12.333,1.2333,15.667 --> 0
( 23.) 1,1,0,0,5000000,28,0,15.833 --> 0
( 24.) 1,1,0,0,208000000,19,0,14 --> 0
( 25.) 3,3,0,0,808000000,16.667,0.5,17 --> 0
( 26.) 1,0,1,43600000,0,8,0,12.583 --> 0
( 27.) 1,0,1,655000000,0,8,0,11.75 --> 0
( 28.) 3,2,2,250.11,259000000,20.333,0,16.833 --> 0
( 29.) 2,1,1,114000000,114000000,14,0.22222,15 --> 0
( 30.) 5,3,2,452000000,457000000,10.4,2.8048,16.833 --> 0
( 31.) 1,1,0,0,235000000,16,0,13.75 --> 0
( 32.) 2,2,0,0,491000000,14,0,14.5 --> 0
( 33.) 1,0,1,142000000,0,16,0,13.833 --> 0
( 34.) 1,1,0,0,4540000,23,0,14.333 --> 0
( 35.) 7,1,6,644000000,228000000,13.286,3.1929,18.667 --> 0
( 36.) 4,3,2,284000000,1190000000,18,0,17 --> 0
( 37.) 1,0,1,304000000,0,18,0,14.167 --> 0
( 38.) 2,2,1,113000000,530000000,19.5,0,15.667 --> 0
( 39.) 6,5,1,903000000,1670000000,15.833,1.7381,18.167 --> 0
( 40.) 1,1,0,0,90300000,7,0,11 --> 0
( 41.) 1,1,0,0,113000000,17,0,14.25 --> 0
( 42.) 1,1,0,0,13800000,21,0,14.583 --> 0
( 43.) 3,3,0,0,394000000,15.333,0,16.333 --> 0
( 44.) 2,2,0,0,980000000,11,0.33333,14.167 --> 0
( 45.) 4,4,0,0,825000000,16.75,0.45833,17.5 --> 0
( 46.) 2,2,0,0,345000000,12,0,14.333 --> 0
( 47.) 4,4,0,0,1430000000,18.75,0.076923,17.167 --> 0
( 48.) 2,2,0,0,508000000,18.5,0,15.833 --> 0
( 49.) 3,3,1,501000000,1110000000,15,0.68182,16.333 --> 0
( 50.) 2,2,0,0,846000000,15.5,0,15.25 --> 0
( 51.) 2,2,0,0,1150000000,22,0,16.667 --> 0
( 52.) 3,3,0,0,543000000,14.667,0.125,16.333 --> 0
( 53.) 2,1,1,198000000,115000000,17,0,16.333 --> 0
( 54.) 2,2,0,0,461000000,16,0.44444,16 --> 0
( 55.) 3,3,1,32000000,356000000,20,0,16.667 --> 0
```

Figura 4.20 Comprobación de resultados en weka. Identificación de días con ruido (anómalos).

CAPITULO 5

ANÁLISIS DE RESULTADOS

5.1 INTRODUCCIÓN

En este capítulo final se presenta de manera gráfica todo los resultados obtenidos del pre procesamiento y procesamiento de los datos manejados en el capítulo anterior, se puede cristalizar el comportamiento que tiene cada uno de los bancos mensualmente así como posibles eventos que durante estos lapsos pudiesen llegar a transcurrir.

A continuación se presenta un análisis físico apoyado de gráficas las cuales representan cada uno de los 38 bancos participantes. Dichas gráficas servirán como un repositorio.

Así mismo se presenta la introducción de un nuevo Dataset más actualizado de los bancos participantes, con la finalidad de que se evalúe el comportamiento y determine si continua siendo normal o ha tenido ciertas variaciones.

5.2 ANÁLISIS DE LOS BANCOS PARTICIPANTES

ANÁLISIS BANCO 1

El análisis mostrado a continuación representa el banco 1, en el cual se detectaron un número de 4 anomalías los días 1, 5, 136 y 284, tales días corresponden a las fechas establecidas en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.2

Banco 1				
Indicadores	Días			
	13-ago-2004	19-ago-2004	21-feb-2005	21-sep-2005
	1	5	136	284
Degree	5.000	17.000	18.000	31.000
DegreeIn	5.000	11.000	18.000	31.000
DegreeOut	0.000	13.000	9.000	30.000
MontosIn	*****			
MontosOut	*****			
Betweenness	9.200	9.882	8.444	15.323
Closeness	1.500	88.950	153.180	0.000
Affinity	13.333	22.000	25.000	31.500

Tabla 5.1 Análisis de anomalías detectadas en banco 1

En las siguientes tablas se presentan la cantidad de clusters obtenidos del banco analizado, en este caso, los resultantes fueron 4, enumerados de '0' al '3', de cada clúster se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster 0		cluster 1	
	Media	desv est	Media	desv est
Degree	20.906	4.123	18.000	4.123
DegreeIn	18.768	4.402	17.000	4.402
DegreeOut	16.051	4.348	11.667	4.348
MontosIn	*****			
MontosOut	*****			
Betweenness	12.931	1.323	8.880	1.323
Closeness	62.854	34.347	87.423	34.347
Affinity	26.106	2.909	24.889	2.909

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

cluster		cluster	
2		3	
Media	desv est	Media	desv est
31.562	0.587	31.143	0.639
31.335	0.727	30.571	0.904
30.394	0.744	30.000	0.926

15.207	0.616	12.655	0.471
39.561	39.767	86.917	41.865
31.766	0.382	31.429	0.416

Tabla 5.2 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 1

ANÁLISIS BANCO 2

El análisis mostrado a continuación representa el banco 2, en el cual se detectaron un número de 4 anomalías los días 68, 80, 82, 156 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 2				
Indicadores	Días			
	17-nov-2004	3-dic-2004	7-dic-2004	22-mar-2005
	68	80	82	156
Degree	27.00	22.00	24.00	1.00
DegreeIn	18.00	20.00	21.00	1.00
DegreeOut	24.00	20.00	19.00	0.00
MontosIn	*****			
MontosOut	*****			
Betweenness	12.67	14.05	14.17	10.00
Closeness	146.72	94.47	88.41	0.00
Affinity	29.50	27.00	28.50	13.17

Tabla 5.3 Análisis de anomalías detectadas en banco 2

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

En las siguientes tablas se presentan la cantidad de clusters obtenidos del banco analizado, en este caso, los resultantes fueron 4, enumerados de '0' al '3', de cada clúster se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster		cluster	
	0		1	
	Media	desv est	Media	desv est
Degree	11.400	1.463	19.649	2.832
DegreeIn	10.150	1.652	17.403	2.871
DegreeOut	6.650	1.558	15.545	3.406
MontosIn	*****			
MontosOut	*****			
Betweenness	10.148	1.137	13.989	1.143
Closeness	35.518	12.562	28.583	22.538
Affinity	19.375	2.025	25.738	1.642

	cluster		cluster	
	2		3	
	Media	desv est	Media	desv est
	0.000	0.000	1.000	0.000
	0.000	0.000	1.000	0.000
	0.000	0.000	0.000	0.000

	0.000	0.000	21.667	5.558
	0.000	0.000	0.000	0.000
	0.000	0.000	15.111	0.926

Tabla 5.4 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 2

ANÁLISIS BANCO 3

El análisis mostrado a continuación representa el banco 3, en el cual se detectaron un número de 8 anomalías los días 1, 17, 41, 64, 74, 111, 136, 203 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 3				
Indicadores	Días			
	13-ago-2004	6-sep-2004	11-oct-2004	11-nov-2004
	1	17	41	64
Degree	9.00	20.00	24.00	24.00
DegreeIn	6.00	14.00	16.00	22.00
DegreeOut	7.00	12.00	17.00	16.00
MontosIn	*****			
MontosOut	*****			
Betweenness	7.56	5.50	7.54	7.21
Closeness	35.18	250.70	284.52	281.18
Affinity	15.67	22.33	26.67	28.33

25-nov-2004	17-enero-2005	21-feb-2005	30-mayo-2005
74	111	136	203
26.00	21.00	22.00	32.00
20.00	16.00	17.00	31.00
17.00	14.00	14.00	32.00

7.69	8.57	7.95	12.25
333.87	129.66	170.78	152.61
29.50	26.17	27.00	32.00

Tabla 5.5 Análisis de anomalías detectadas en banco 3

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

En las siguientes tablas se presentan la cantidad de clusters obtenidos del banco analizado, en este caso, los resultantes fueron 3, enumerados de '0' al '2', de cada clúster se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster		cluster	
	0		1	
	Media	desv est	Media	desv est
Degree	14.333	2.181	26.361	1.655
DegreeIn	12.400	1.781	24.197	2.063
DegreeOut	10.867	2.579	20.754	2.546
MontosIn	*****			
MontosOut	*****			
Betweenness	10.123	0.976	12.367	1.338
Closeness	52.464	39.758	105.410	63.291
Affinity	20.789	2.023	29.141	0.948

cluster	
2	
Media	desv est
31.357	0.578
30.852	0.806
30.933	0.714

15.278	0.700
31.393	23.477
31.662	0.348

Tabla 5.6 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 3

ANÁLISIS BANCO 4

El análisis mostrado a continuación representa el banco 4, en el cual se detectó una anomalía el día 13, tal día corresponde a la fecha establecida en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 4	
Indicadores	Días
	31-ago-2004
	13
Degree	16.00
DegreeIn	9.00
DegreeOut	11.00
MontosIn	*****
MontosOut	*****
Betweenness	10.00
Closeness	155.34
Affinity	23.50

Tabla 5.7 Análisis de anomalías detectadas en banco 4

En las siguientes tablas se presentan la cantidad de clusters obtenidos del banco analizado, en este caso el resultante fue 1, con una numeración de '0', del cual se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster	
	0	
	Media	desv est
Degree	25.674	6.279
DegreeIn	23.877	7.279
DegreeOut	22.149	7.306
MontosIn	*****	
MontosOut	*****	
Betweenness	14.792	1.731
Closeness	32.539	31.550
Affinity	28.665	3.583

Tabla 5.8 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 4

ANÁLISIS BANCO 5

El análisis mostrado a continuación representa el banco 5, en el cual se detectaron un número de 2 anomalías los días 115 y 145 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 5		
Indicadores	Días	
	2-sep-2004	4-mar-2005
	15	145
Degree	17.00	31.00
DegreeIn	12.00	28.00
DegreeOut	12.00	25.00
MontosIn	*****	
MontosOut	*****	
Betweenness	9.41	14.13
Closeness	159.12	141.26
Affinity	24.00	32.00

Tabla 5.9 Análisis de anomalías detectadas en banco 5

En las siguientes tablas se presentan la cantidad de clusters obtenidos del banco analizado, en este caso, los resultantes fueron 2, enumerados de '0' al '1', de cada clúster se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster		cluster	
	0		1	
	Media	desv est	Media	desv est
Degree	15.217	2.266	28.910	0.893
DegreeIn	13.420	2.465	27.595	1.339
DegreeOut	9.231	2.361	25.748	1.483
MontosIn	*****			
MontosOut	*****			
Betweenness	14.818	2.250	16.071	0.748
Closeness	13.626	9.602	20.040	11.365
Affinity	23.167	1.780	30.429	0.455

Tabla 5.10 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 5

ANÁLISIS BANCO 6

El análisis mostrado a continuación representa el banco 6, en el cual se detectaron un número de 3 anomalías los días 68, 80, 82 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 6			
Indicadores	Días		
	17-nov-2004	3-dic-2004	7-dic-2004
	68	80	82
Degree	2.00	5.00	4.00
DegreeIn	2.00	2.00	3.00
DegreeOut	0.00	5.00	3.00
MontosIn	*****		
MontosOut	*****		
Betweenness	11.00	20.80	27.75
Closeness	1.75	0.00	0.00
Affinity	15.00	17.00	18.00

Tabla 5.11 Análisis de anomalías detectadas en banco 6

En las siguientes tablas se presentan la cantidad de clusters obtenidos del banco analizado, en este caso, los resultantes fueron 2, enumerados de '0' al '1', de cada clúster se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster		cluster	
	0		1	
	Media	desv est	Media	desv est
Degree	0.000	0.000	4.554	1.837
DegreeIn	0.000	0.000	3.580	1.639
DegreeOut	0.000	0.000	3.619	1.682
MontosIn	*****			
MontosOut	*****			
Betweenness	0.000	0.000	26.142	4.324
Closeness	0.000	0.000	0.002	0.018
Affinity	0.000	0.000	17.850	1.743

Tabla 5.12 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 6

ANÁLISIS BANCO 7

El análisis mostrado a continuación representa el banco 7, en el cual se detectaron un número de 8 anomalías los días 1, 2, 6, 17, 28, 64, 69, 92 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días

Banco 7				
Indicadores	Días			
	13-ago-2004	16-ago-2004	20-ago-2004	6-sep-2004
	1	2	6	17
Degree	1.00	5.00	5.00	2.00
DegreeIn	0.00	3.00	3.00	0.00
DegreeOut	1.00	3.00	3.00	2.00
MontosIn	*****			
MontosOut	*****			
Betweenness	3.00	8.00	9.40	14.50
Closeness	0.00	55.36	51.17	0.00
Affinity	8.12	13.00	15.00	12.75

22-sep-2004	11-nov-2004	18-nov-2004	21-dic-2004
28	64	69	92
8.00	12.00	12.00	21.00
7.00	6.00	10.00	14.00
4.00	8.00	10.00	16.00

13.13	9.92	13.92	13.10
72.12	108.80	134.03	135.03
18.83	22.50	21.83	26.50

Tabla 5.13 Análisis de anomalías detectadas en banco 7

En las siguientes tablas se presentan la cantidad de clusters obtenidos del banco analizado, en este caso el resultante fue 1, con una numeración de '0', del cual se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster	
	0	
	Media	desv est
Degree	16.991	3.871
DegreeIn	13.922	4.238
DegreeOut	13.210	3.861
MontosIn	*****	
MontosOut	*****	
Betweenness	17.500	3.102
Closeness	13.189	16.349
Affinity	24.336	2.338

Tabla 5.14 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 7

ANÁLISIS BANCO 8

El análisis mostrado a continuación representa el banco 8, en el cual se detectaron un número de 10 anomalías los días 1, 17, 23, 64, 74, 111, 146, 167, 179, 193 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 8					
Indicadores	Días				
	13-ago-2004	6-sep-2004	14-sep-2004	11-nov-2004	25-nov-2004
	1	17	23	64	74
Degree	5.00	1.00	13.00	3.00	5.00
DegreeIn	2.00	0.00	7.00	2.00	5.00
DegreeOut	4.00	1.00	10.00	1.00	0.00
MontosIn	*****				
MontosOut	*****				
Betweenness	6.20	9.00	11.23	14.33	13.00
Closeness	19.13	0.00	133.65	13.65	5.94
Affinity	13.33	10.17	22.00	17.33	18.17

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

17-enero-2005	7-mar-2005	8-abr-2005	26-abr-2005	16-mayo-2005
111	146	167	179	193
6.00	22.00	22.00	23.00	20.00
2.00	18.00	18.00	17.00	14.00
4.00	18.00	15.00	21.00	17.00

9.50	18.73	18.23	17.91	19.00
64.08	80.03	73.49	80.16	75.73
18.33	27.50	27.50	28.00	26.50

Tabla 5.15 Análisis de anomalías detectadas en banco 8

En las siguientes tablas se presentan la cantidad de clusters obtenidos del banco analizado, en este caso el resultante fue 1, con una numeración de '0', del cual se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster	
	0	
	Media	desv est
Degree	18.154	3.853
DegreeIn	14.522	3.986
DegreeOut	13.704	3.887
MontosIn	*****	
MontosOut	*****	
Betweenness	17.219	2.683
Closeness	12.995	14.159
Affinity	24.903	2.271

Tabla 5.16 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 8

ANÁLISIS BANCO 9

El análisis mostrado a continuación representa el banco 9, en el cual se detectaron un número de 11 anomalías los días 1, 7, 8, 16, 24, 78, 88, 111, 128, 284, 348 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 9					
Indicadores	Días				
	13-ago-2004	23-ago-2004	24-ago-2004	3-sep-2004	15-sep-2004
	1	7	8	16	24
Degree	11.000	4.000	14.000	10.000	4.000
DegreeIn	6.000	2.000	5.000	5.000	3.000
DegreeOut	7.000	4.000	12.000	9.000	2.000
MontosIn	*****				
MontosOut	*****				
Betweenness	7.000	8.250	9.786	10.400	20.000
Closeness	71.997	10.488	65.317	76.861	0.000
Affinity	16.667	14.167	21.167	19.667	17.833

1-dic-2004	15-dic-2004	17-enero-2005	9-feb-2005	21-sep-2005	21-dic-2005
78	88	111	128	284	348
25.000	13.000	8.000	16.000	18.000	20.000
6.000	11.000	4.000	10.000	17.000	19.000
25.000	11.000	4.000	12.000	14.000	17.000

12.960	18.923	12.375	15.563	20.000	19.050
154.840	0.946	73.367	81.469	5.533	9.416
29.000	22.833	19.000	24.000	25.000	26.000

Tabla 5.17 Análisis de anomalías detectadas en banco 9

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

En la siguiente tabla se presenta la cantidad de clusters obtenidos del banco analizado, en este caso el resultante fue 1, con una numeración de '0', del cual se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster	
	0	
	Media	desv est
Degree	18.122	4.391
DegreeIn	14.660	4.766
DegreeOut	13.610	3.349
MontosIn	*****	
MontosOut	*****	
Betweenness	17.567	2.024
Closeness	10.203	5.938
Affinity	24.878	2.568

Tabla 5.18 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 9

ANÁLISIS BANCO 10

El análisis mostrado a continuación representa el banco 10, en el cual se detectaron un número de 12 anomalías los días 14, 15, 74, 122, 128, 149, 150, 182, 283, 293, 325, 355 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

Asimismo se muestran los valores resultantes de los indicadores en los presentes

Banco 10				
Indicadores	Días			
	1-sep-2004	2-sep-2004	25-nov-2004	1-feb-2005
	14	15	74	122
Degree	5.000	3.000	4.000	9.000
DegreeIn	3.000	3.000	0.000	4.000
DegreeOut	2.000	0.000	4.000	7.000
MontosIn	*****			
MontosOut	*****			
Betweenness	8.400	11.000	7.000	18.444
Closeness	20.542	11.583	8.867	6.535
Affinity	15.833	15.167	16.833	20.333
9-feb-2005	10-mar-2005	11-mar-2005	29-abr-2005	20-sep-2005
128	149	150	182	283
8.000	17.000	16.000	13.000	10.000
3.000	13.000	7.000	9.000	9.000
7.000	13.000	15.000	12.000	9.000

16.125	20.294	20.750	23.000	25.300
5.927	5.522	4.459	1.639	0.100
19.833	24.500	24.000	22.500	21.000
4-oct-2005	17-nov-2005	30-dic-2005		
293	325	355		
10.000	16.000	7.000		
8.000	7.000	7.000		
7.000	16.000	6.000		

24.300	21.063	28.000		
0.100	1.394	0.000		
21.000	24.000	19.500		

Tabla 5.19 Análisis de anomalías detectadas en banco 10

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

En la siguiente tabla se presenta la cantidad de clusters obtenidos del banco analizado, en este caso los resultantes fueron 2, con una numeración de '0' y '1', del cual se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster		cluster	
	0		1	
	Media	desv est	Media	desv est
Degree	3.173	1.631	11.480	1.601
DegreeIn	2.367	1.368	9.118	1.297
DegreeOut	1.144	1.142	9.451	1.628
MontosIn	*****			
MontosOut	*****			
Betweenness	16.406	4.837	24.032	1.409
Closeness	0.791	1.766	0.648	0.682
Affinity	16.073	2.286	21.718	0.827

Tabla 5.20 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 10

ANÁLISIS BANCO 11

El análisis mostrado a continuación representa el banco 11, en el cual se detectaron un número de 8 anomalías los días 8, 9, 14, 27, 74, 154, 155, 323 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 11				
Indicadores	Días			
	24-ago-2004	25-ago-2004	1-sep-2004	21-sep-2004
	8	9	14	27
Degree	5.000	6.000	8.000	8.000
DegreeIn	4.000	2.000	2.000	3.000
DegreeOut	2.000	6.000	8.000	5.000
MontosIn	*****			
MontosOut	*****			
Betweenness	4.400	8.667	8.250	10.625
Closeness	78.731	80.399	56.866	80.853
Affinity	15.083	16.500	17.667	20.000

25-nov-2004	17-mar-2005	18-mar-2005	15-nov-2005
74	154	155	323
6.000	18.000	16.000	15.000
1.000	18.000	11.000	8.000
5.000	12.000	13.000	13.000

10.333	21.333	20.188	22.333
84.000	3.075	6.196	1.110
19.167	24.500	24.000	23.500

Tabla 5.21 Análisis de anomalías detectadas en banco 11

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

En la siguiente tabla se presenta la cantidad de clusters obtenidos del banco analizado, en este caso los resultantes fueron 2, con una numeración de '0' y '1', del cual se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster		cluster	
	0		1	
	Media	desv est	Media	desv est
Degree	12.713	3.580	11.167	1.344
DegreeIn	8.144	4.383	3.667	1.106
DegreeOut	10.026	2.670	10.000	1.155
MontosIn	*****			
MontosOut	*****			
Betweenness	19.310	3.392	14.925	1.487
Closeness	6.060	7.479	88.625	10.107
Affinity	22.140	2.294	21.389	1.153

Tabla 5.22 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 11

ANÁLISIS BANCO 12

El análisis mostrado a continuación representa el banco 12, en el cual se detectaron un número de 6 anomalías los días 7, 12, 111, 126, 136, 355 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 12			
Indicadores	Días		
	23-ago-2004	30-ago-2004	17-enero-2005
	7	12	111
Degree	12.000	14.000	17.000
DegreeIn	5.000	7.000	10.000
DegreeOut	10.000	11.000	13.000
MontosIn	*****		
MontosOut	*****		
Betweenness	7.917	9.643	9.412
Closeness	117.290	199.570	105.770
Affinity	19.167	21.500	24.000

Banco 12		
7-feb-2005	21-feb-2005	30-dic-2005
126	136	355
23.000	18.000	28.000
17.000	13.000	26.000
20.000	11.000	23.000

12.478	8.667	15.321
196.630	127.000	49.858
27.500	24.667	30.000

Tabla 5.23 Análisis de anomalías detectadas en banco 12

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

En la siguiente tabla se presenta la cantidad de clusters obtenidos del banco analizado, en este caso los resultantes fueron 2, con una numeración de '0' y '1', del cual se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster		cluster	
	0		1	
	Media	desv est	Media	desv est
Degree	19.050	4.145	28.029	1.352
DegreeIn	14.482	3.733	26.352	1.765
DegreeOut	13.612	4.099	23.395	1.616
MontosIn	*****			
MontosOut	*****			
Betweenness	13.448	1.577	16.347	0.751
Closeness	49.799	34.953	37.980	11.643
Affinity	25.152	2.879	29.987	0.697

Tabla 5.24 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 12

ANÁLISIS BANCO 13

El análisis mostrado a continuación representa el banco 13, en el cual se detectaron un número de 6 anomalías los días 92, 07, 102, 104, 108, 353 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 13			
Indicadores	Días		
	21-dic-2004	28-dic-2004	4-enero-2005
	92	97	102
Degree	12.000	11.000	7.000
DegreeIn	12.000	10.000	6.000
DegreeOut	5.000	4.000	1.000
MontosIn	*****		
MontosOut	*****		
Betweenness	17.000	12.818	13.286
Closeness	24.999	163.450	85.652
Affinity	21.833	21.833	18.833

6-enero-2005	12-enero-2005	28-dic-2005
104	108	353
15.000	15.000	13.000
14.000	12.000	12.000
11.000	10.000	10.000

15.333	15.067	21.385
92.834	94.826	1.643
23.333	23.333	22.500

Tabla 5.25 Análisis de anomalías detectadas en banco 13

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

En la siguiente tabla se presenta la cantidad de clusters obtenidos del banco analizado, en este caso los resultantes fueron 2, con una numeración de '0' y '1', del cual se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster		cluster	
	0		1	
	Media	desv est	Media	desv est
Degree	0.000	0.000	14.733	4.468
DegreeIn	0.000	0.000	12.608	4.366
DegreeOut	0.000	0.000	10.084	4.437
MontosIn	*****			
MontosOut	*****			
Betweenness	0.000	0.000	18.520	3.121
Closeness	0.000	0.000	7.832	9.892
Affinity	0.000	0.000	23.146	2.707

Tabla 5.26 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 13

ANÁLISIS BANCO 14

El análisis mostrado a continuación representa el banco 14, en el cual se detectaron un número de 16 anomalías los días 1,2, 6, 9, 17, 18, 19, 27, 32, 70, 71, 111, 112, 113, 205, 263 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 14					
Indicadores	Días				
	13-ago-2004	16-ago-2004	20-ago-2004	25-ago-2004	6-sep-2004
	1	2	6	9	17
Degree	3.000	1.000	1.000	7.000	0.000
DegreeIn	3.000	1.000	0.000	4.000	0.000
DegreeOut	0.000	1.000	1.000	6.000	0.000
MontosIn	*****				
MontosOut	*****				
Betweenness	6.667	5.000	5.000	10.571	0.000
Closeness	48.702	0.000	0.000	86.961	0.000
Affinity	11.917	8.417	9.833	17.167	0.000

7-sep-2004	8-sep-2004	21-sep-2004	28-sep-2004	19-nov-2004
18	19	27	32	70
9.000	6.000	7.000	16.000	9.000
3.000	6.000	3.000	6.000	8.000
8.000	5.000	6.000	13.000	6.000

14.333	15.833	12.000	11.125	17.889
4.140	0.600	97.641	177.500	95.013
19.500	18.000	19.500	23.833	20.667

22-nov-2004	17-enero-2005	18-enero-2005	19-enero-2005	1-jun-2005	22-ago-2005
71	111	112	113	205	263
7.000	2.000	9.000	9.000	14.000	11.000
6.000	1.000	7.000	6.000	12.000	11.000
2.000	1.000	3.000	6.000	12.000	0.000

15.571	4.000	15.667	16.111	23.500	23.000
83.692	5.118	95.458	17.438	1.429	0.786
19.167	12.450	20.333	20.500	23.000	21.500

Tabla 5.27 Análisis de anomalías detectadas en banco 14

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

En la siguiente tabla se presenta la cantidad de clusters obtenidos del banco analizado, en este caso el resultante fue 1, con una numeración de '0', del cual se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster	
	0	
	Media	desv est
Degree	12.558	3.718
DegreeIn	9.796	3.642
DegreeOut	9.106	3.353
MontosIn	*****	
MontosOut	*****	
Betweenness	19.768	3.349
Closeness	4.380	5.956
Affinity	22.029	2.329

Tabla 5.28 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 14

ANÁLISIS BANCO 15

El análisis mostrado a continuación representa el banco 15 en el cual se detectaron un número de 11 anomalías los días 1, 2, 8, 17, 20, 25, 26, 136, 137, 340, 350 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 15					
Indicadores	Días				
	13-ago-2004	16-ago-2004	24-ago-2004	6-sep-2004	9-sep-2004
	1	2	8	17	20
Degree	0.000	0.000	6.000	0.000	12.000
DegreeIn	0.000	0.000	3.000	0.000	3.000
DegreeOut	0.000	0.000	4.000	0.000	9.000
MontosIn	*****				
MontosOut	*****				
Betweenness	0.000	0.000	9.167	0.000	11.167
Closeness	0.000	0.000	69.008	0.000	55.749
Affinity	0.000	0.000	16.417	0.000	20.833

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

17-sep-2004	20-sep-2004	21-feb-2005	22-feb-2005	8-dic-2005	23-dic-2005
25	26	136	137	340	350
5.000	8.000	4.000	10.000	14.000	13.000
2.000	1.000	3.000	4.000	9.000	8.000
3.000	7.000	2.000	10.000	13.000	10.000

12.000	12.750	16.250	19.200	22.714	22.000
80.625	99.517	0.277	1.204	2.358	1.891
17.833	19.667	17.333	20.333	23.000	22.500

Tabla 5.29 Análisis de anomalías detectadas en banco 15

En la siguiente tabla se presenta la cantidad de clusters obtenidos del banco analizado, en este caso el resultante fue 1, con una numeración de '0', del cual se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster	
	0	
	Media	desv est
Degree	10.448	3.157
DegreeIn	7.308	3.133
DegreeOut	8.160	2.775
MontosIn	*****	
MontosOut	*****	
Betweenness	20.576	4.006
Closeness	2.283	3.499
Affinity	20.991	1.938

Tabla 5.30 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 15

ANÁLISIS BANCO 16

El análisis mostrado a continuación representa el banco 16, en el cual se detectaron un número de 9 anomalías los días 1, 2, 7, 9, 13, 14, 15, 17, 34 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 16				
Indicadores	Días			
	13-ago-2004	16-ago-2004	23-ago-2004	25-ago-2004
	1	2	7	9
Degree	0.000	0.000	6.000	7.000
DegreeIn	0.000	0.000	4.000	6.000
DegreeOut	0.000	0.000	2.000	4.000
MontosIn	*****			
MontosOut	*****			
Betweenness	0.000	0.000	7.500	9.143
Closeness	0.000	0.000	57.749	77.378
Affinity	0.000	0.000	15.667	16.833

31-ago-2004	1-sep-2004	2-sep-2004	6-sep-2004	30-sep-2004
13	14	15	17	34
11.000	9.000	9.000	9.000	8.000
3.000	7.000	3.000	3.000	2.000
10.000	5.000	6.000	6.000	7.000

9.455	9.222	9.889	9.000	14.125
107.170	79.724	95.879	72.879	16.366
20.167	18.000	19.333	16.667	19.667

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

1-oct-2004	11-oct-2004	1-nov-2004	1-dic-2004
35	41	56	78
9.000	7.000	6.000	10.000
5.000	1.000	6.000	7.000
4.000	6.000	0.000	6.000

14.000	6.714	12.667	14.900
23.254	121.260	3.625	100.130
19.333	17.167	18.167	21.333

Tabla 5.31 Análisis de anomalías detectadas en banco 16

En la siguiente tabla se presenta la cantidad de clusters obtenidos del banco analizado, en este caso el resultante fue 1, con una numeración de '0', del cual se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster	
	0	
	Media	desv est
Degree	11.023	3.529
DegreeIn	8.696	4.108
DegreeOut	8.503	3.070
MontosIn	*****	
MontosOut	*****	
Betweenness	20.221	3.307
Closeness	2.238	3.720
Affinity	21.300	2.079

Tabla 5.32 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 16

ANÁLISIS BANCO 17

El análisis mostrado a continuación representa el banco 17, en el cual se detectaron un número de 7 anomalías los días 1, 12, 25, 64, 95, 96, 228 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 17			
Indicadores	Días		
	13-ago-2004	30-ago-2004	17-sep-2004
	1	12	25
Degree	5.000	20.000	19.000
DegreeIn	4.000	10.000	13.000
DegreeOut	4.000	20.000	12.000
MontosIn	*****		
MontosOut	*****		
Betweenness	9.600	7.500	10.737
Closeness	0.400	272.280	239.380
Affinity	12.917	24.333	25.000
11-nov-2004	24-dic-2004	27-dic-2004	4-jul-2005
64	95	96	228
17.000	22.000	22.000	26.000
11.000	19.000	14.000	23.000
15.000	12.000	22.000	23.000

8.765	12.773	12.000	13.885
322.810	71.050	105.630	54.132
25.000	27.000	27.500	29.000

Tabla 5.33 Análisis de anomalías detectadas en banco 17

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

En la siguiente tabla se presenta la cantidad de clusters obtenidos del banco analizado, en este caso los resultantes fueron 2, con una numeración de '0' y '1', del cual se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster		cluster	
	0		1	
	Media	desv est	Media	desv est
Degree	18.978	3.789	28.167	1.031
DegreeIn	16.580	3.409	27.271	1.226
DegreeOut	11.152	5.188	22.743	1.762
MontosIn	*****			
MontosOut	*****			
Betweenness	13.546	1.853	16.328	0.672
Closeness	54.567	46.528	45.363	8.259
Affinity	25.121	2.657	30.067	0.535

Tabla 5.34 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 17

ANÁLISIS BANCO 18

El análisis mostrado a continuación representa el banco 18, en el cual se detectó una anomalía el día 13, tal día corresponde a la fecha establecida en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 18		
Indicadores	Días	
	27-oct-2004	7-enero-2005
	53	105
Degree	15.00	10.00
DegreeIn	12.00	8.00
DegreeOut	13.00	10.00
MontosIn	*****	
MontosOut	*****	
Betweenness	17.20	17.90
Closeness	5.04	6.44
Affinity	23.50	20.83

Tabla 5.35 Análisis de anomalías detectadas en banco 18

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

En las siguientes tablas se presentan la cantidad de clusters obtenidos del banco analizado, en este caso, los resultantes fueron 4, enumerados de '0' al '3', de cada clúster se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster		cluster	
	0		1	
	Media	desv est	Media	desv est
Degree	0.000	0.000	4.886	2.327
DegreeIn	0.000	0.000	4.600	2.393
DegreeOut	0.000	0.000	1.429	1.460
MontosIn	*****			
MontosOut	*****			
Betweenness	0.000	0.000	16.646	2.456
Closeness	0.000	0.000	0.361	0.632
Affinity	0.000	0.000	16.905	2.401

	cluster		cluster	
	2		3	
	Media	desv est	Media	desv est
	10.646	1.326	13.125	0.927
	9.063	1.435	10.000	1.000
	9.550	1.256	11.875	0.781

	22.751	2.704	17.234	0.588
	0.655	0.728	6.904	1.585
	21.283	0.694	22.542	0.832

Tabla 5.36 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 18

ANÁLISIS BANCO 19

El análisis mostrado a continuación representa el banco 19, en el cual se detectaron un número de 2 anomalías los días 115 y 145 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 19				
Indicadores	Días			
	14-sep-2004	18-nov-2004	28-mar-2005	31-mar-2005
	23	69	158	161
Degree	3.000	2.000	9.000	10.000
DegreeIn	2.000	1.000	8.000	8.000
DegreeOut	1.000	2.000	8.000	9.000
MontosIn	*****			
MontosOut	*****			
Betweenness	6.333	6.500	24.778	23.100
Closeness	59.620	55.537	0.467	0.804
Affinity	14.500	13.750	20.500	21.000

13-dic-2005	19-dic-2005	26-dic-2005
342	346	351
9.000	12.000	15.000
6.000	8.000	9.000
9.000	9.000	11.000

26.111	23.000	19.667
0.222	0.717	3.583
20.500	22.000	23.500

Tabla 5.37 Análisis de anomalías detectadas en banco 19

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

En las siguientes tablas se presentan la cantidad de clusters obtenidos del banco analizado, en este caso, los resultantes fueron 3, enumerados de '0' al '2', de cada clúster se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster		cluster		cluster	
	0		1		2	
	Media	desv est	Media	desv est	Media	desv est
Degree	0.000	0.000	2.643	1.212	9.454	1.412
DegreeIn	0.000	0.000	1.512	1.175	7.029	1.350
DegreeOut	0.000	0.000	1.465	1.162	7.483	1.542
MontosIn	*****					
MontosOut	*****					
Betweenness	0.000	0.000	15.349	5.175	25.175	1.466
Closeness	0.000	0.000	1.058	2.669	0.323	0.565
Affinity	0.000	0.000	15.693	1.782	20.705	0.710

Tabla 5.38 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 19

ANÁLISIS BANCO 20

El análisis mostrado a continuación representa el banco 20, en el cual se detectaron un número de 8 anomalías los días 17, 26, 27, 48, 97, 215, 284, 348 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 20				
Indicadores	Días			
	6-sep-2004	20-sep-2004	21-sep-2004	20-oct-2004
	17	26	27	48
Degree	1.000	12.000	12.000	20.000
DegreeIn	1.000	8.000	9.000	17.000
DegreeOut	0.000	8.000	7.000	18.000
MontosIn	*****			
MontosOut	*****			
Betweenness	20.000	14.167	13.083	15.200
Closeness	0.000	19.341	20.107	27.487
Affinity	12.083	22.000	22.333	26.000
	28-dic-2004	15-jun-2005	21-sep-2005	21-dic-2005
	97	215	284	348
	14.000	20.000	18.000	15.000
	9.000	17.000	16.000	13.000
	10.000	14.000	17.000	15.000

	13.643	18.750	20.667	21.867
	16.102	11.398	2.595	1.994
	23.167	26.000	25.000	23.500

Tabla 5.39 Análisis de anomalías detectadas en banco 20

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

En las siguientes tablas se presentan la cantidad de clusters obtenidos del banco analizado, en este caso, los resultantes fueron 2, enumerados de '0' al '1', de cada clúster se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster		cluster	
	0		1	
	Media	desv est	Media	desv est
Degree	0.000	0.000	16.435	2.605
DegreeIn	0.000	0.000	13.766	2.678
DegreeOut	0.000	0.000	14.631	3.041
MontosIn	*****			
MontosOut	*****			
Betweenness	0.000	0.000	18.992	2.463
Closeness	0.000	0.000	5.595	3.643
Affinity	0.000	0.000	24.148	1.521

Tabla 5.40 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 20

ANÁLISIS BANCO 21

El análisis mostrado a continuación representa el banco 21, en el cual se detectaron un número de 10 anomalías los días 4, 5, 8, 9, 10, 15, 49, 50, 85, 143 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 21				
Indicadores	Días			
	18-ago-2004 4	19-ago-2004 5	24-ago-2004 8	25-ago-2004 9
Degree	4.00	4.00	1.00	4.00
DegreeIn	4.00	2.00	1.00	2.00
DegreeOut	1.00	4.00	1.00	3.00
MontosIn	*****			
MontosOut	*****			
Betweenness	12.50	15.00	11.00	14.25
Closeness	0.24	0.00	0.00	0.00
Affinity	13.83	14.83	11.67	15.00

26-ago-2004 10	2-sep-2004 15	21-oct-2004 49	22-oct-2004 50	10-dic-2004 85	2-mar-2005 143
0.00	3.00	7.00	8.00	8.00	9.00
0.00	1.00	6.00	7.00	3.00	4.00
0.00	2.00	6.00	7.00	8.00	8.00

0.00	16.00	20.86	19.88	18.50	16.00
0.00	0.00	0.00	0.07	46.23	102.29
0.00	15.50	19.00	19.50	20.33	20.50

Tabla 5.41 Análisis de anomalías detectadas en banco 21

En la siguiente tabla se presenta la cantidad de clusters obtenidos del banco analizado, en este caso el resultante fue 1, con una numeración de '0', del cual se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

	cluster	
	0	
	Media	desv est
Degree	7.957	2.057
DegreeIn	5.777	2.188
DegreeOut	6.528	1.808
MontosIn	*****	
MontosOut	*****	
Betweenness	23.065	3.566
Closeness	0.310	1.228
Affinity	19.713	1.600

Tabla 5.42 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 21

ANÁLISIS BANCO 22

El análisis mostrado a continuación representa el banco 22, en el cual se detectaron un número de 6 anomalías los días 1, 3, 4, 17, 101, 117 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 22			
Indicadores	Días		
	13-ago-2004	17-ago-2004	18-ago-2004
	1	3	4
Degree	14.000	12.000	10.000
DegreeIn	6.000	10.000	8.000
DegreeOut	11.000	6.000	6.000
MontosIn	*****		
MontosOut	*****		
Betweenness	6.000	9.167	9.800
Closeness	193.880	81.123	64.218
Affinity	18.333	18.333	17.333

6-sep-2004			3-enero-2005			25-enero-2005		
17			101			117		
4.000			12.000			18.000		
1.000			6.000			18.000		
3.000			9.000			10.000		

8.250			14.667			15.389		
47.033			11.127			110.080		
12.833			21.833			25.000		

Tabla 5.43 Análisis de anomalías detectadas en banco 22

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

En las siguientes tablas se presentan la cantidad de clusters obtenidos del banco analizado, en este caso, los resultantes fueron 4, enumerados de '0' al '3', de cada clúster se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster		cluster	
	0		1	
	Media	desv est	Media	desv est
Degree	14.385	1.862	17.799	1.856
DegreeIn	8.692	1.976	14.941	2.493
DegreeOut	11.000	2.038	14.623	2.312
MontosIn	*****			
MontosOut	*****			
Betweenness	9.446	0.906	18.634	2.676
Closeness	116.820	19.230	11.984	11.327
Affinity	20.987	1.597	24.870	0.980

cluster		cluster	
2		3	
Media	desv est	Media	desv est
9.000	1.095	9.714	1.485
4.000	0.632	7.429	1.400
8.000	1.414	8.143	0.990

13.360	1.568	22.810	1.781
16.968	12.495	1.337	1.015
20.100	0.841	20.714	0.881

Tabla 5.44 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 22

ANÁLISIS BANCO 23

El análisis mostrado a continuación representa el banco 23, en el cual se detectaron un número de 4 anomalías los días 115, 127, 336, 337 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 23				
Indicadores	Días			
	21-enero-2005	8-feb-2005	2-dic-2005	5-dic-2005
	115	127	336	337
Degree	2.00	11.00	8.00	6.00
DegreeIn	2.00	7.00	6.00	6.00
DegreeOut	0.00	8.00	8.00	6.00
MontosIn	*****			
MontosOut	*****			
Betweenness	14.00	17.91	25.50	25.83
Closeness	60.80	102.46	0.00	0.17
Affinity	16.33	21.00	20.00	19.00

Tabla 5.45 Análisis de anomalías detectadas en banco 23

En la siguiente tabla se presenta la cantidad de clusters obtenidos del banco analizado, en este caso el resultante fue 1, con una numeración de '0', del cual se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

	cluster		cluster	
	0		1	
	Media	desv est	Media	desv est
Degree	6.135	2.661	0.000	0.000
DegreeIn	3.950	1.880	0.000	0.000
DegreeOut	4.863	2.841	0.000	0.000
MontosIn	*****			
MontosOut	*****			
Betweenness	23.465	4.725	0.000	0.000
Closeness	0.434	1.889	0.000	0.000
Affinity	18.619	2.119	0.000	0.000

Tabla 5.46 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 23

ANÁLISIS BANCO 24

El análisis mostrado a continuación representa el banco 24, en el cual se detectaron un número de 3 anomalías los días 111, 307, 331 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 24			
Indicadores	Días		
	17-enero-2005	24-oct-2005	25-nov-2005
	111	307	331
Degree	3.00	9.00	10.00
DegreeIn	2.00	6.00	7.00
DegreeOut	1.00	8.00	10.00
MontosIn	*****		
MontosOut	*****		
Betweenness	7.33	26.67	26.30
Closeness	54.19	0.00	0.07
Affinity	15.67	20.50	21.00

Tabla 5.47 Análisis de anomalías detectadas en banco 24

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

En la siguiente tabla se presentan la cantidad de clusters obtenidos del banco analizado, en este caso, los resultantes fueron 3, enumerados de '0' al '2', de cada clúster se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster		cluster		cluster	
	0		1		2	
	Media	desv est	Media	desv est	Media	desv est
Degree	1.704	0.813	0.000	0.000	9.632	1.673
DegreeIn	1.183	0.654	0.000	0.000	7.110	1.557
DegreeOut	0.635	0.816	0.000	0.000	8.000	1.761
MontosIn	*****					
MontosOut	*****					
Betweenness	16.563	6.404	0.000	0.000	25.128	1.497
Closeness	0.577	1.317	0.000	0.000	0.434	0.527
Affinity	14.915	1.541	0.000	0.000	20.794	0.840

Tabla 5.48 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 24

ANÁLISIS BANCO 25

El análisis mostrado a continuación representa el banco 25, en el cual se detectaron un número de 12 anomalías los días 17, 38, 45, 46, 48, 50, 53, 55 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 25				
Indicadores	Días			
	6-sep-2004	6-oct-2004	15-oct-2004	18-oct-2004
	17	38	45	46
Degree	1.000	6.000	6.000	6.000
DegreeIn	0.000	5.000	4.000	2.000
DegreeOut	1.000	1.000	5.000	6.000
MontosIn	*****			
MontosOut	*****			
Betweenness	4.000	20.667	21.167	22.833
Closeness	0.000	0.000	0.000	0.000
Affinity	8.667	18.667	18.167	18.833

20-oct-2004	22-oct-2004	27-oct-2004	29-oct-2004
48	50	53	55
7.000	5.000	8.000	6.000
3.000	4.000	5.000	5.000
7.000	5.000	6.000	5.000

20.857	21.600	20.375	20.500
0.071	0.000	0.000	0.182
19.167	17.833	19.667	18.833

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

4-nov-2004	5-nov-2004	10-nov-2004	12-nov-2004
59	60	63	65
6.000	6.000	5.000	4.000
5.000	2.000	4.000	3.000
4.000	6.000	5.000	4.000

22.333	22.500	20.400	22.750
0.000	0.000	0.000	0.000
18.667	18.333	18.333	18.000

Tabla 5.49 Análisis de anomalías detectadas en banco 25

En la siguiente tabla se presentan la cantidad de clusters obtenidos del banco analizado, en este caso, los resultantes fueron 3, enumerados de '0' al '2', de cada clúster se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos

	cluster		cluster		cluster	
	0		1		2	
	Media	desv est	Media	desv est	Media	desv est
Degree	2.271	1.132	5.000	0.000	0.000	0.000
DegreeIn	1.417	1.133	4.000	0.632	0.000	0.000
DegreeOut	1.396	0.699	4.000	0.632	0.000	0.000
MontosIn	*****					
MontosOut	*****					
Betweenness	18.903	4.015	21.920	0.466	0.000	0.000
Closeness	0.000	0.000	0.000	0.000	0.000	0.000
Affinity	14.955	2.358	18.167	0.105	0.000	0.000

Tabla 5.50 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 25

ANÁLISIS BANCO 26

El análisis mostrado a continuación representa el banco 26, en el cual se detectaron un número de 19 anomalías los días 1, 2, 6, 9, 11, 12, 16, 17, 18, 25, 26, 29, 30, 31, 32, 33, 34, 47, 306 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 26				
Indicadores	Días			
	13-ago-2004	16-ago-2004	20-ago-2004	25-ago-2004
	1	2	6	9
Degree	4.000	10.000	13.000	11.000
DegreeIn	4.000	3.000	4.000	5.000
DegreeOut	2.000	9.000	13.000	10.000
MontosIn	*****			
MontosOut	*****			
Betweenness	10.750	9.200	9.615	11.818
Closeness	0.000	25.268	87.547	84.611
Affinity	12.750	15.417	19.833	19.333

27-ago-2004	30-ago-2004	3-sep-2004	6-sep-2004	7-sep-2004
11	12	16	17	18
13.000	9.000	13.000	2.000	15.000
11.000	3.000	9.000	2.000	8.000
8.000	8.000	10.000	2.000	15.000

11.000	10.222	10.538	16.000	13.800
18.038	7.814	39.503	0.000	18.715
19.833	18.167	21.500	12.750	23.000

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

17-sep-2004	20-sep-2004	23-sep-2004	24-sep-2004	27-sep-2004
25	26	29	30	31
15.000	12.000	12.000	17.000	13.000
9.000	8.000	9.000	8.000	7.000
14.000	8.000	10.000	17.000	12.000

12.733	13.500	15.167	13.706	15.769
104.750	109.120	5.873	42.707	7.214
22.833	21.833	21.500	24.500	22.000

28-sep-2004	29-sep-2004	30-sep-2004	19-oct-2004	21-oct-2005
32	33	34	47	306
11.000	12.000	12.000	17.000	12.000
8.000	11.000	8.000	9.000	12.000
10.000	7.000	11.000	17.000	8.000

16.909	16.750	17.583	16.647	23.333
1.593	2.622	0.313	7.189	0.827
21.500	20.833	22.000	24.000	22.000

Tabla 5.51 Análisis de anomalías detectadas en banco 26

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

En la siguiente tabla se presenta la cantidad de clusters obtenidos del banco analizado, en este caso el resultante fue 1, con una numeración de '0', del cual se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster	
	0	
	Media	desv est
Degree	11.887	2.246
DegreeIn	8.902	2.322
DegreeOut	9.708	2.203
MontosIn	*****	
MontosOut	*****	
Betweenness	21.014	3.588
Closeness	3.916	7.178
Affinity	21.811	1.277

Tabla 5.52 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 26

ANÁLISIS BANCO 27

El análisis mostrado a continuación representa el banco 27, en el cual se detectó una anomalía el día 76, tal día corresponde a la fecha establecida en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 27	
Indicadores	Días
	29-nov-2004
	76
Degree	1.00
DegreeIn	1.00
DegreeOut	0.00
MontosIn	*****
MontosOut	*****
Betweenness	11.00
Closeness	0.00
Affinity	13.25

Tabla 5.53 Análisis de anomalías detectadas en banco 27

En la siguiente tabla se presenta la cantidad de clusters obtenidos del banco analizado, en este caso el resultante fue 1, con una numeración de '0', del cual se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster	
	0	
	Media	desv est
Degree	0.000	0.000
DegreeIn	0.000	0.000
DegreeOut	0.000	0.000
MontosIn	*****	
MontosOut	*****	
Betweenness	0.000	0.000
Closeness	0.000	0.000
Affinity	0.000	0.000

Tabla 5.54 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 27

ANÁLISIS BANCO 28

El análisis mostrado a continuación representa el banco 28, en el cual se detectaron un número de 14 anomalías los días 1, 10, 22, 34, 48, 62, 65, 111, 136, 146, 147, 153, 297, 352 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 28				
Indicadores	Días			
	13-ago-2004	26-ago-2004	13-sep-2004	30-sep-2004
	1	10	22	34
Degree	3.000	4.000	8.000	9.000
DegreeIn	1.000	4.000	5.000	8.000
DegreeOut	3.000	1.000	8.000	5.000
MontosIn	*****			
MontosOut	*****			
Betweenness	10.333	13.500	14.750	16.778
Closeness	0.000	0.000	3.239	2.116
Affinity	11.750	14.167	18.500	20.500
20-oct-2004	9-nov-2004	12-nov-2004	17-enero-2005	21-feb-2005
48	62	65	111	136
11.000	11.000	11.000	1.000	2.000
10.000	11.000	10.000	1.000	1.000
11.000	11.000	11.000	0.000	1.000

19.455	17.727	19.727	8.000	12.500
0.992	5.413	0.682	0.000	0.000
21.333	22.000	21.833	12.000	14.833

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

7-mar-2005	8-mar-2005	16-mar-2005	10-oct-2005	27-dic-2005
146	147	153	297	352
9.000	8.000	10.000	0.000	11.000
8.000	8.000	7.000	0.000	3.000
9.000	5.000	8.000	0.000	11.000

25.111	20.875	23.400	0.000	22.091
0.000	0.250	0.321	0.000	1.521
21.000	20.000	21.000	0.000	21.500

Tabla 5.55 Análisis de anomalías detectadas en banco 28

En la siguiente tabla se presenta la cantidad de clusters obtenidos del banco analizado, en este caso el resultante fue 1, con una numeración de '0', del cual se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster	
	0	
	Media	desv est
Degree	8.595	1.977
DegreeIn	7.208	2.074
DegreeOut	7.164	2.199
MontosIn	*****	
MontosOut	*****	
Betweenness	21.902	3.527
Closeness	0.559	0.701
Affinity	20.043	1.540

Tabla 5.56 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 28

ANÁLISIS BANCO 29

El análisis mostrado a continuación representa el banco 29, en el cual se detectaron un número de 10 anomalías los días 24, 71, 155, 292, 294, 314, 315, 323, 324 ,335 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 29					
Indicadores	Días				
	15-sep-2004	22-nov-2004	18-mar-2005	3-oct-2005	5-oct-2005
	24	71	155	292	294
Degree	13.00	3.00	3.00	3.00	3.00
DegreeIn	1.00	3.00	1.00	3.00	3.00
DegreeOut	13.00	2.00	3.00	1.00	3.00
MontosIn	*****				
MontosOut	*****				
Betweenness	13.31	23.00	27.67	29.67	27.67
Closeness	82.29	0.00	0.00	0.00	0.00
Affinity	22.50	16.83	17.50	17.50	17.50

2-nov-2005	3-nov-2005	15-nov-2005	16-nov-2005	1-dic-2005
314	315	323	324	335
1.00	2.00	3.00	1.00	3.00
1.00	1.00	2.00	1.00	3.00
1.00	2.00	3.00	1.00	3.00

31.00	32.00	28.00	31.00	24.00
0.00	0.00	0.00	0.00	0.00
16.33	17.00	17.50	16.33	17.50

Tabla 5.57 Análisis de anomalías detectadas en banco 29

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

En las siguientes tablas se presentan la cantidad de clusters obtenidos del banco analizado, en este caso, los resultantes fueron 4, enumerados de '0' al '3', de cada clúster se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster		cluster	
	0		1	
	Media	desv est	Media	desv est
Degree	0.000	0.000	2.286	0.589
DegreeIn	0.000	0.000	2.000	0.000
DegreeOut	0.000	0.000	1.607	0.838
MontosIn	*****			
MontosOut	*****			
Betweenness	0.000	0.000	30.068	2.963
Closeness	0.000	0.000	0.000	0.000
Affinity	0.000	0.000	17.101	0.330

cluster		cluster	
2		3	
Media	desv est	Media	desv est
1.316	0.494	1.087	0.282
1.000	0.000	0.000	0.000
1.065	0.607	1.087	0.282

27.839	4.520	29.196	3.767
0.000	0.000	0.000	0.000
16.088	0.822	16.083	0.691

Tabla 5.58 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 29

ANÁLISIS BANCO 30

El análisis mostrado a continuación representa el banco 30, en el cual se detectaron un número de 12 anomalías los días 1, 2, 14, 119, 120, 121, 188, 214, 265, 352, 353, 354 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 30				
Indicadores	Días			
	13-ago-2004	16-ago-2004	1-sep-2004	27-enero-2005
	1	2	14	119
Degree	0.000	0.000	6.000	7.000
DegreeIn	0.000	0.000	3.000	6.000
DegreeOut	0.000	0.000	5.000	4.000
MontosIn	*****			
MontosOut	*****			
Betweenness	0.000	0.000	10.333	21.143
Closeness	0.000	0.000	25.842	0.250
Affinity	0.000	0.000	15.833	19.500

28-enero-2005	31-enero-2005	9-mayo-2005	14-jun-2005
120	121	188	214
6.000	8.000	14.000	12.000
4.000	2.000	13.000	9.000
3.000	7.000	12.000	11.000

22.667	18.875	22.357	23.250
0.000	0.846	2.603	1.046
18.833	19.833	23.000	22.000

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

24-ago-2005	27-dic-2005	28-dic-2005	29-dic-2005
265	352	353	354
14.000	17.000	17.000	13.000
14.000	14.000	8.000	8.000
7.000	11.000	15.000	12.000

20.429	20.471	18.529	22.923
4.710	6.410	13.187	1.154
23.000	24.500	24.500	22.500

Tabla 5.59 Análisis de anomalías detectadas en banco 30

En la siguiente tabla se presenta la cantidad de clusters obtenidos del banco analizado, en este caso el resultante fue 1, con una numeración de '0', del cual se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster	
	0	
	Media	desv est
Degree	9.286	3.254
DegreeIn	6.376	3.308
DegreeOut	7.726	3.150
MontosIn	*****	
MontosOut	*****	
Betweenness	21.816	3.662
Closeness	1.095	1.576
Affinity	20.367	2.059

Tabla 5.60 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 30

ANÁLISIS BANCO 33

El análisis mostrado a continuación representa el banco 31, en el cual se detectaron un número de 8 anomalías los días 20, 58, 93, 190, 191, 221, 237, 240 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 33				
Indicadores	Días			
	9-sep-2004	3-nov-2004	22-dic-2004	11-mayo-2005
	20	58	93	190
Degree	7.000	11.000	17.000	16.000
DegreeIn	6.000	5.000	9.000	12.000
DegreeOut	3.000	8.000	13.000	13.000
MontosIn	*****			
MontosOut	*****			
Betweenness	14.286	16.091	16.000	21.563
Closeness	94.672	117.480	57.327	2.498
Affinity	18.333	21.500	25.000	24.000
	12-mayo-2005	23-jun-2005	15-jul-2005	20-jul-2005
	191	221	237	240
	18.000	14.000	12.000	14.000
	14.000	10.000	8.000	12.000
	15.000	14.000	10.000	9.000

	21.778	20.786	22.333	22.857
	4.666	4.353	1.849	1.752
	25.000	23.000	22.000	23.000

Tabla 5.61 Análisis de anomalías detectadas en banco 31

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

En las siguientes tablas se presentan la cantidad de clusters obtenidos del banco analizado, en este caso, los resultantes fueron 2, enumerados de '0' al '1', de cada clúster se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster		cluster	
	0		1	
	Media	desv est	Media	desv est
Degree	0.000	0.000	11.036	3.812
DegreeIn	0.000	0.000	7.152	3.226
DegreeOut	0.000	0.000	8.429	3.334
MontosIn	*****			
MontosOut	*****			
Betweenness	0.000	0.000	20.487	3.471
Closeness	0.000	0.000	4.066	6.410
Affinity	0.000	0.000	21.351	2.211

Tabla 5.62 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 31

ANÁLISIS BANCO 34

El análisis mostrado a continuación representa el banco 34, en el cual se detectaron un número de 14 anomalías los días 3, 12, 17, 76, 156, 158, 182, 191, 205, 224, 227, 312, 323, 334 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 34				
Indicadores	Días			
	17-ago-2004	30-ago-2004	6-sep-2004	29-nov-2004
	3	12	17	76
Degree	4.000	2.000	0.000	11.000
DegreeIn	1.000	2.000	0.000	5.000
DegreeOut	3.000	0.000	0.000	9.000
MontosIn	*****			
MontosOut	*****			
Betweenness	9.000	7.500	0.000	16.273
Closeness	56.667	53.286	0.000	117.660
Affinity	13.667	13.167	0.000	21.833

22-mar-2005	28-mar-2005	29-abr-2005	12-mayo-2005	1-jun-2005
156	158	182	191	205
10.000	10.000	9.000	17.000	7.000
4.000	6.000	4.000	9.000	5.000
9.000	9.000	7.000	15.000	6.000

21.800	25.300	24.000	21.588	24.000
134.440	0.133	0.919	5.406	6.000
21.500	21.000	20.500	24.500	19.500

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

28-jun-2005	1-jul-2005	31-oct-2005	15-nov-2005	30-nov-2005
224	227	312	323	334
7.000	5.000	12.000	12.000	9.000
7.000	2.000	6.000	9.000	4.000
4.000	5.000	9.000	9.000	8.000

22.000	27.400	22.750	23.750	25.444
0.432	0.000	143.470	1.687	0.286
19.500	18.500	22.500	22.000	20.500

Tabla 5.63 Análisis de anomalías detectadas en banco 34

En la siguiente tabla se presenta la cantidad de clusters obtenidos del banco analizado, en este caso el resultante fue 1, con una numeración de '0', del cual se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster	
	0	
	Media	desv est
Degree	8.965	2.839
DegreeIn	5.176	2.436
DegreeOut	7.255	2.713
MontosIn	*****	
MontosOut	*****	
Betweenness	21.583	4.333
Closeness	1.908	3.339
Affinity	20.217	1.904

Tabla 5.64 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 34

ANÁLISIS BANCO 35

El análisis mostrado a continuación representa el banco 35, en el cual se detectaron un número de 10 anomalías los días 6, 9, 49, 50, 102, 114, 122, 287, 341, 344 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 35				
Indicadores	Días			
	20-ago-2004	25-ago-2004	21-oct-2004	22-oct-2004
	6	9	49	50
Degree	0.00	0.00	5.00	4.00
DegreeIn	0.00	0.00	2.00	2.00
DegreeOut	0.00	0.00	3.00	3.00
MontosIn	*****			
MontosOut	*****			
Betweenness	0.00	0.00	19.40	13.75
Closeness	0.00	0.00	0.14	1.24
Affinity	0.00	0.00	17.83	16.67

4-enero-2005	20-enero-2005	1-feb-2005	26-sep-2005	9-dic-2005	15-dic-2005
102	114	122	287	341	344
7.00	3.00	5.00	10.00	7.00	8.00
3.00	2.00	0.00	6.00	4.00	5.00
5.00	3.00	5.00	10.00	7.00	7.00

16.57	19.67	17.40	25.20	28.86	27.13
8.07	0.22	15.49	0.69	0.00	0.15
18.67	16.83	18.00	21.00	19.50	20.00

Tabla 5.65 Análisis de anomalías detectadas en banco 35

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

En la siguiente tabla se presenta la cantidad de clusters obtenidos del banco analizado, en este caso el resultante fue 1, con una numeración de '0', del cual se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster	
	0	
	Media	desv est
Degree	6.461	2.445
DegreeIn	3.872	2.293
DegreeOut	5.159	2.026
MontosIn	*****	
MontosOut	*****	
Betweenness	22.886	4.728
Closeness	0.380	0.689
Affinity	18.798	2.068

Tabla 5.66 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 35

ANÁLISIS BANCO 36

El análisis mostrado a continuación representa el banco 36, en el cual se detectaron un número de 14 anomalías los días 15, 16, 17, 24, 25, 26, 27, 29, 54, 61, 97, 113, 132, 352 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 36				
Indicadores	Días			
	2-sep-2004	3-sep-2004	6-sep-2004	15-sep-2004
	15	16	17	24
Degree	15.000	13.000	4.000	8.000
DegreeIn	8.000	5.000	2.000	7.000
DegreeOut	13.000	13.000	3.000	2.000
MontosIn	*****			
MontosOut	*****			
Betweenness	8.867	10.231	10.750	13.125
Closeness	134.230	140.460	1.200	7.394
Affinity	22.667	21.167	13.833	20.000

17-sep-2004	20-sep-2004	21-sep-2004	23-sep-2004	28-oct-2004
25	26	27	29	54
11.000	14.000	15.000	15.000	25.000
5.000	10.000	10.000	12.000	12.000
10.000	11.000	9.000	10.000	24.000

14.182	12.643	11.133	12.200	12.520
9.401	128.850	133.310	130.120	218.900
20.833	22.833	23.667	23.000	29.000

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

8-nov-2004	28-dic-2004	19-enero-2005	15-feb-2005	27-dic-2005
61	97	113	132	352
19.000	12.000	23.000	12.000	27.000
15.000	6.000	21.000	10.000	16.000
15.000	9.000	14.000	11.000	20.000

12.684	12.500	12.174	12.917	16.074
254.090	116.190	126.460	127.670	77.250
26.000	22.167	27.500	22.000	29.500

Tabla 5.67 Análisis de anomalías detectadas en banco 36

En la siguiente tabla se presenta la cantidad de clusters obtenidos del banco analizado, en este caso el resultante fue 1, con una numeración de '0', del cual se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster	
	0	
	Media	desv est
Degree	20.012	3.916
DegreeIn	14.572	3.644
DegreeOut	16.534	3.976
MontosIn	*****	
MontosOut	*****	
Betweennes	16.296	2.484
Closeness	36.400	30.112
Affinity	25.826	2.408

Tabla 5.68 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 36

ANÁLISIS BANCO 37

El análisis mostrado a continuación representa el banco 37, en el cual se detectaron un número de 16 anomalías los días 1, 2, 10, 15, 17, 32, 127, 135, 137, 140, 181, 193, 226, 270, 287, 350 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 37				
Indicadores	Días			
	13-ago-2004	16-ago-2004	26-ago-2004	2-sep-2004
	1	2	10	15
Degree	3.000	4.000	4.000	6.000
DegreeIn	1.000	3.000	1.000	2.000
DegreeOut	3.000	1.000	4.000	4.000
MontosIn	*****			
MontosOut	*****			
Betweenness	5.000	4.750	11.250	8.833
Closeness	48.875	51.340	4.744	72.029
Affinity	11.417	11.667	14.833	17.667

6-sep-2004	28-sep-2004	8-feb-2005	18-feb-2005
17	32	127	135
4.000	6.000	0.000	0.000
0.000	1.000	0.000	0.000
4.000	6.000	0.000	0.000

10.500	18.667	0.000	0.000
62.167	0.000	0.000	0.000
14.000	18.500	0.000	0.000

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

22-feb-2005	25-feb-2005	28-abr-2005	16-mayo-2005
137	140	181	193
0.000	0.000	5.000	9.000
0.000	0.000	3.000	6.000
0.000	0.000	4.000	7.000

0.000	0.000	23.400	25.889
0.000	0.000	0.000	0.382
0.000	0.000	18.500	20.833

30-jun-2005	31-ago-2005	26-sep-2005	23-dic-2005
226	270	287	350
6.000	8.000	6.000	6.000
5.000	7.000	3.000	5.000
6.000	6.000	6.000	5.000

27.833	28.500	29.000	26.167
0.000	0.000	0.000	0.000
19.000	20.000	19.000	19.000

Tabla 5.69 Análisis de anomalías detectadas en banco 37

En la siguiente tabla se presenta la cantidad de clusters obtenidos del banco analizado, en este caso el resultante fue 1, con una numeración de '0', del cual se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

	cluster	
	0	
	Media	desv est
Degree	6.569	1.825
DegreeIn	3.670	2.271
DegreeOut	5.121	1.578
MontosIn	*****	
MontosOut	*****	
Betweenness	22.872	4.929
Closeness	0.967	2.660
Affinity	18.998	1.503

Tabla 5.70 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 37

ANÁLISIS BANCO 38

El análisis mostrado a continuación representa el banco 38, en el cual se detectaron un número de 13 anomalías los días 9, 27, 11, 136, 19, 201, 205, 261, 321, 325, 340, 349, 355 tales días corresponden a las fechas establecidas en la parte superior de la tabla.

Asimismo se muestran los valores resultantes de los indicadores en los presentes días.

Banco 38					
Indicadores	Días				
	25-ago-2004	21-sep-2004	27-ago-2004	21-feb-2005	
	9	27	11	136	
Degree	1.000	1.000	2.000	6.000	
DegreeIn	1.000	1.000	0.000	0.000	
DegreeOut	1.000	1.000	2.000	6.000	
MontosIn	*****				
MontosOut	*****				
Betweenness	7.000	7.000	3.000	10.333	
Closeness	0.000	0.000	4.167	6.189	
Affinity	10.917	12.417	11.700	18.000	

	20-mayo-2005	26-mayo-2005	1-jun-2005	18-ago-2005	11-nov-2005
	197	201	205	261	321
	9.000	5.000	7.000	10.000	8.000
	3.000	3.000	2.000	2.000	3.000
	9.000	4.000	7.000	10.000	7.000

	25.333	25.000	24.714	23.600	21.125
	0.000	0.000	0.000	0.300	0.654
	20.500	18.333	19.333	21.000	19.667

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

17-nov-2005	8-dic-2005	22-dic-2005	30-dic-2005
325	340	349	355
11.000	8.000	10.000	14.000
3.000	3.000	2.000	2.000
10.000	8.000	9.000	13.000

21.818	23.625	25.400	21.571
0.682	0.336	0.000	1.711
21.333	20.000	21.000	23.000

Tabla 5.71 Análisis de anomalías detectadas en banco 38

En las siguientes tablas se presentan la cantidad de clusters obtenidos del banco analizado, en este caso, los resultantes fueron 4, enumerados de '0' al '3', de cada clúster se calculó la media y la desviación estándar de cada uno de los indicadores. Con base en los resultados obtenidos se pueden determinar los rangos, los cuales estos indicadores tomaron en cuenta para determinar dichos días anómalos.

	cluster 0		cluster 1		cluster 2		cluster 3	
	Media	desv est	Media	desv est	Media	desv est	Media	desv est
Degree	0.000	0.000	3.826	1.956	7.724	1.618	7.970	1.640
DegreeIn	0.000	0.000	0.000	0.000	1.000	0.000	2.000	0.000
DegreeOut	0.000	0.000	3.826	1.956	7.724	1.618	7.620	1.725
MontosIn	*****							
MontosOut	*****							
Betweenness	0.000	0.000	14.573	3.574	23.713	1.758	24.081	1.148
Closeness	0.000	0.000	0.424	0.453	0.215	0.307	0.203	0.217
Affinity	0.000	0.000	16.345	2.005	19.778	0.906	19.937	0.871

Tabla 5.72 Análisis estadístico de la media y desviación estándar de los indicadores de red en banco 38

5.3 CONCLUSIONES Y TRABAJOS FUTUROS

El objetivo fundamental de esta tesis como se mencionó en el capítulo 1, era el análisis de los indicadores de red obtenidos del SPEI para así poder ejecutar un algoritmo el cual permitiera identificar los comportamientos de los bancos y con base en esto poder detectar anomalías (outliers), las cuales aportan una solución ligada a la toma de decisiones. Es importante destacar que para el preprocesamiento y procesamiento de los datos es de mucha importancia la selección de los atributos y en este caso de los indicadores de red sobre los cuales se trabajó en esta tesis, dicha selección determinará el funcionamiento adecuado o el fracaso durante la implementación de los algoritmos. Los resultados obtenidos durante este trabajo se muestran de una manera específica basada en una base de datos histórica seleccionada por el banco central, asistido por la implementación de un algoritmo llamado DBSCAN con el cual se logró eliminar el ruido existente de los datos, obteniendo de los archivos mensuales los días anómalos, cabe mencionar que no precisamente por ser un día anómalo quiere decir que sea malo, si no que por algún motivo se comportó de manera distinta al resto de los demás días.

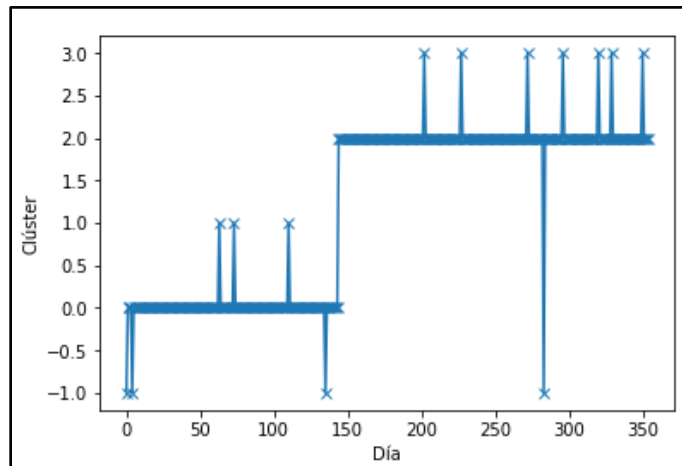
Dichos resultado obtenido pueden utilizarse en la realización de trabajos futuros breves en un periodo corto:

- Realizar pruebas con otros algoritmos Kmeans, COBWEB.
- Analizar resultados tomando un almacén de datos con un periodo de datos más extenso, para así poder identificar más patrones y obtener resultados más precisos.
- Implementar otras técnicas, para poder realizar la selección de los atributos de una manera equitativa, en la que un atributo no sea más dominante que otro.

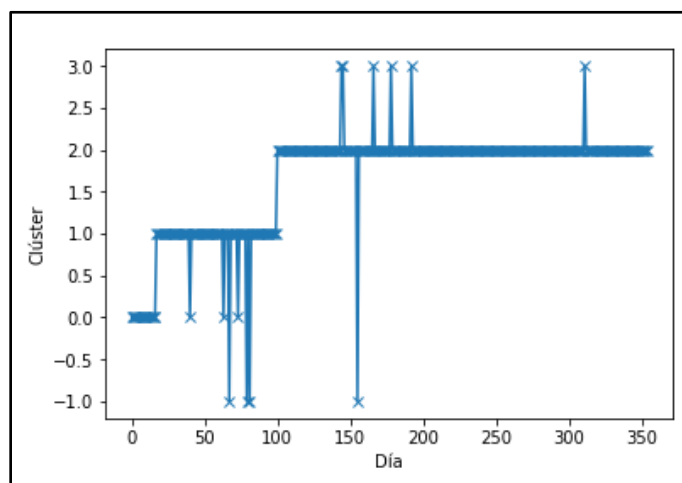
ANEXO 1

GRÁFICAS OBTENIDAS AL EJECUTAR EL ALGORITMO DBSCAN EN PYTHON MOSTRANDO LOS DÍAS ANÓMALOS DE CADA BANCO.

BANCO 1

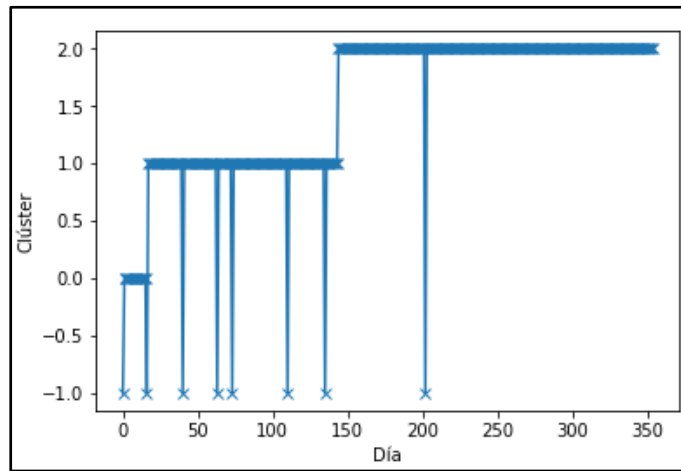


BANCO 2

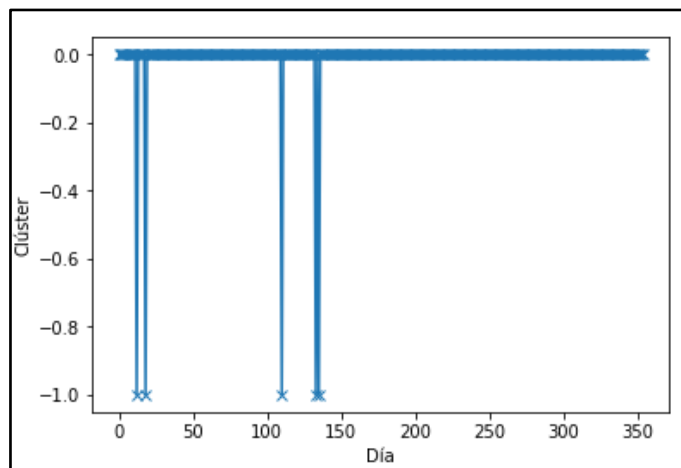


IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

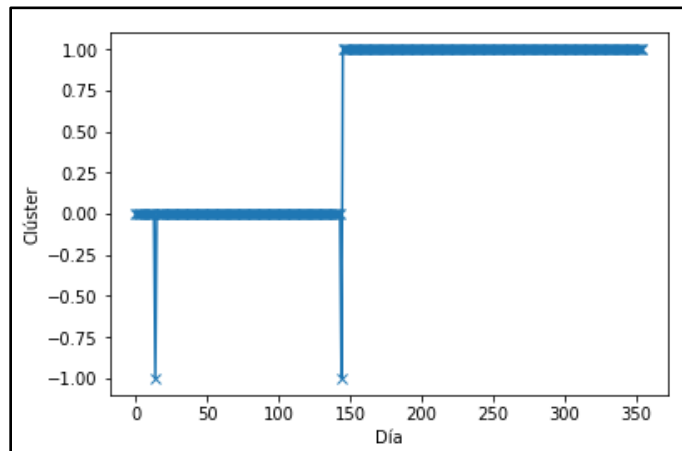
BANCO 3



BANCO 4

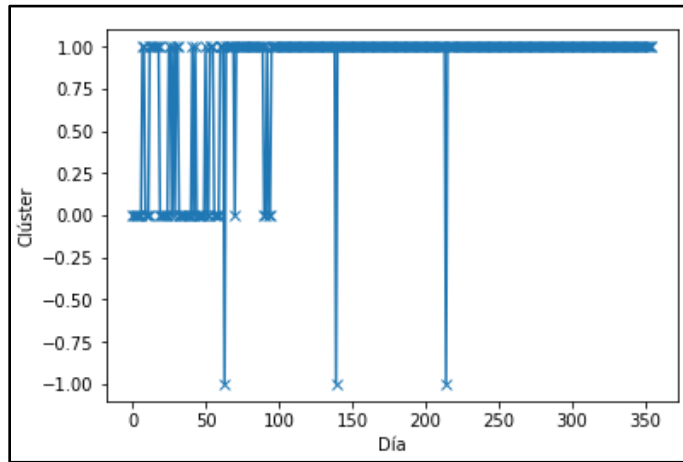


BANCO 5

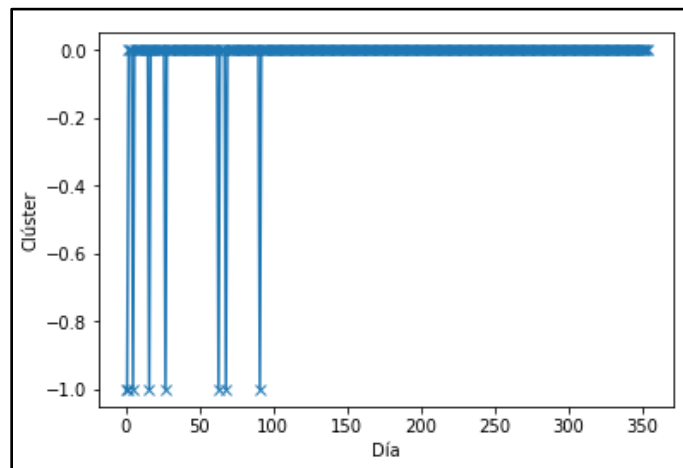


IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

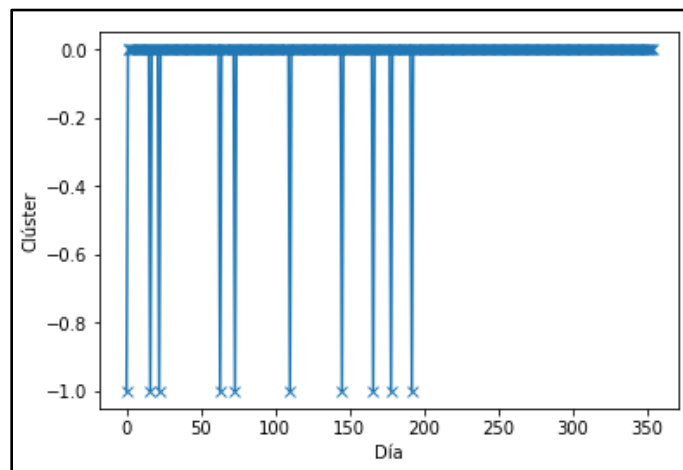
BANCO 6



BANCO 7

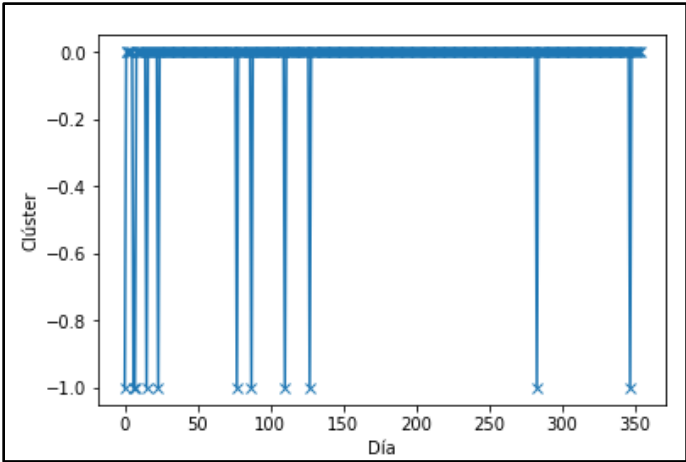


BANCO 8

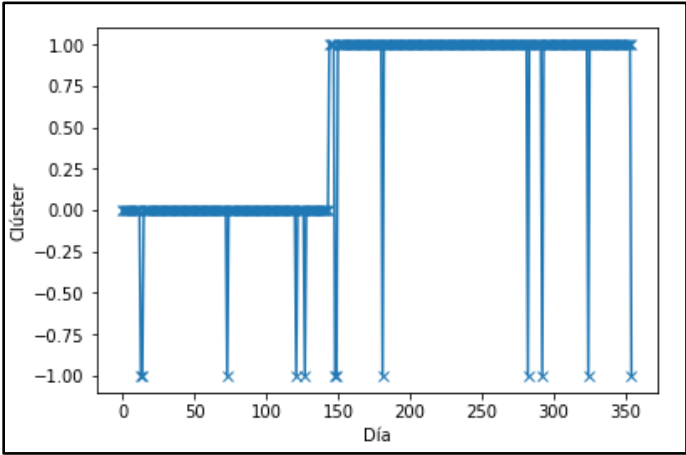


IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

BANCO 9

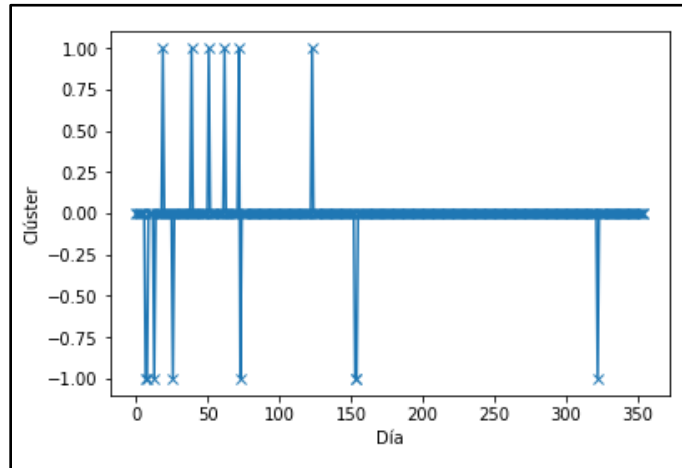


BANCO 10

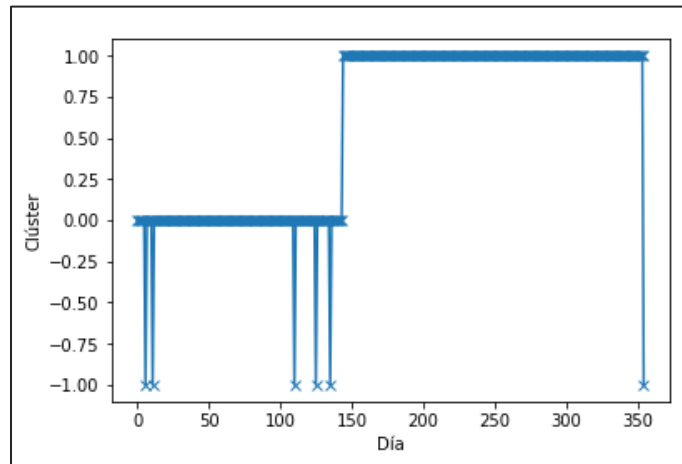


IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

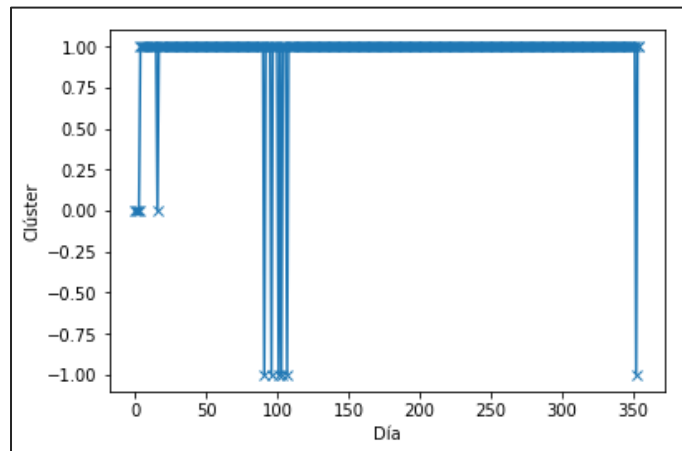
BANCO 11



BANCO 12

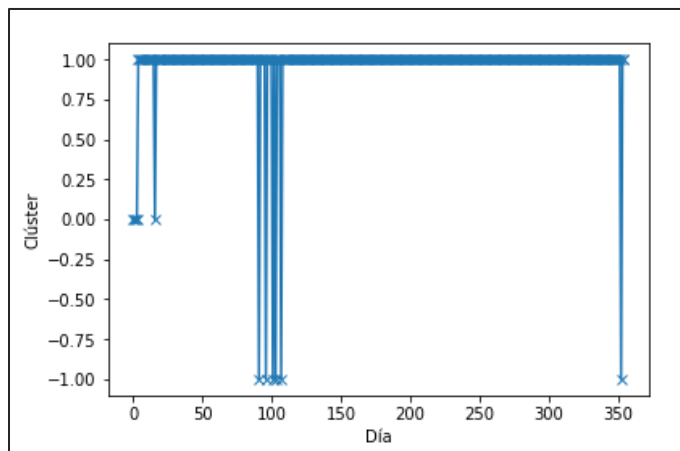


BANCO 13

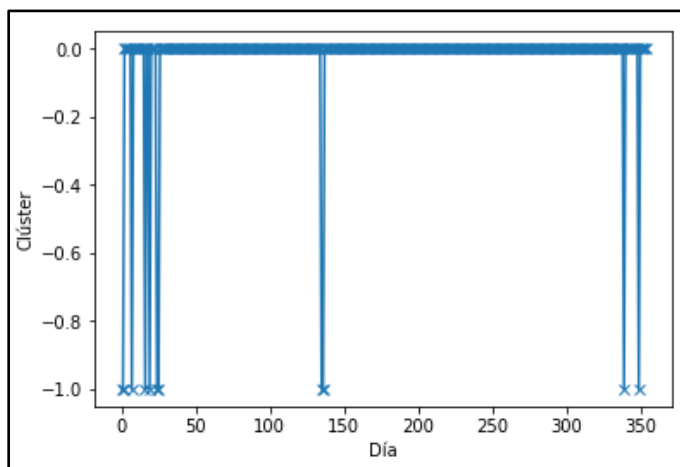


IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

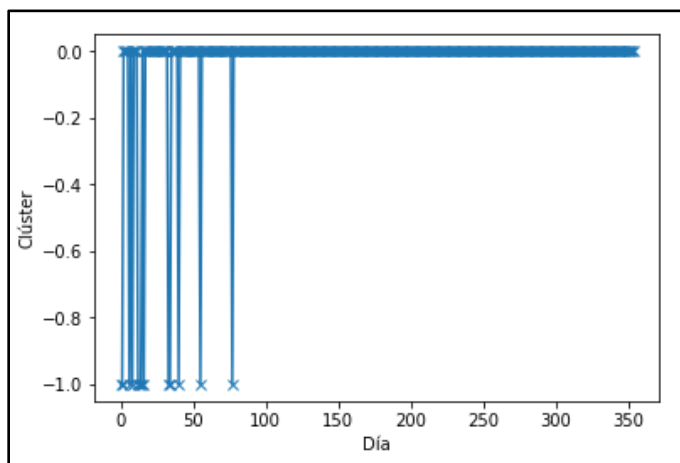
BANCO 14



BANCO 15

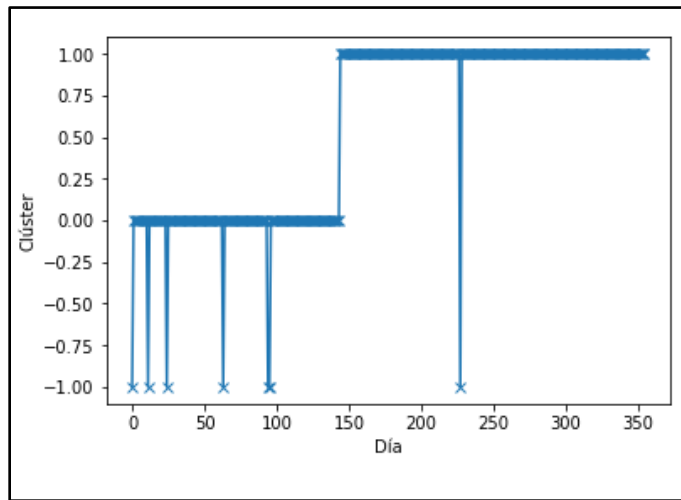


BANCO 16

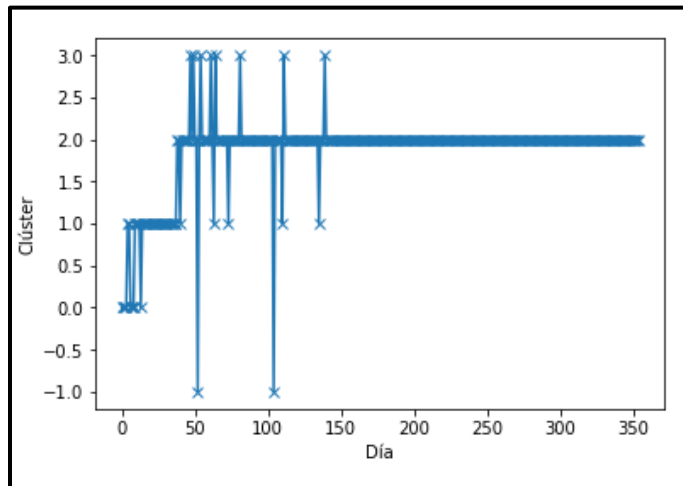


IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

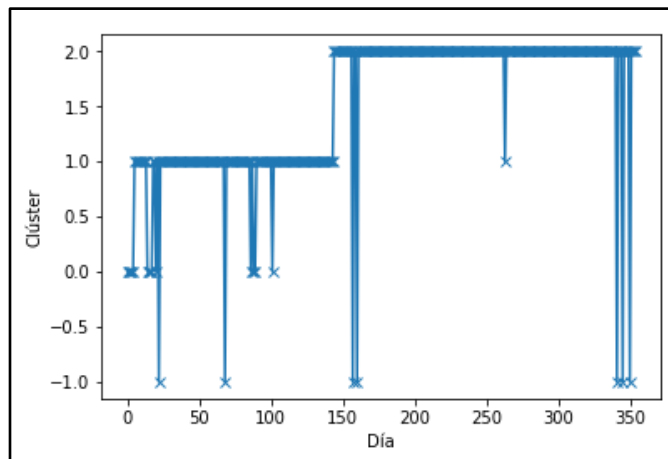
BANCO 17



BANCO 18

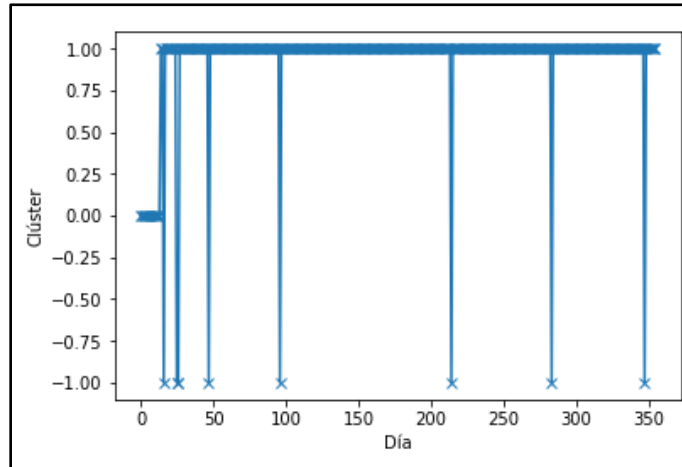


BANCO 19

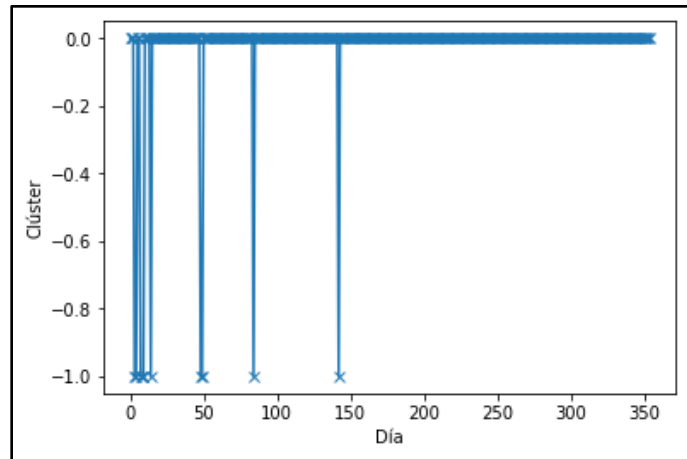


IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

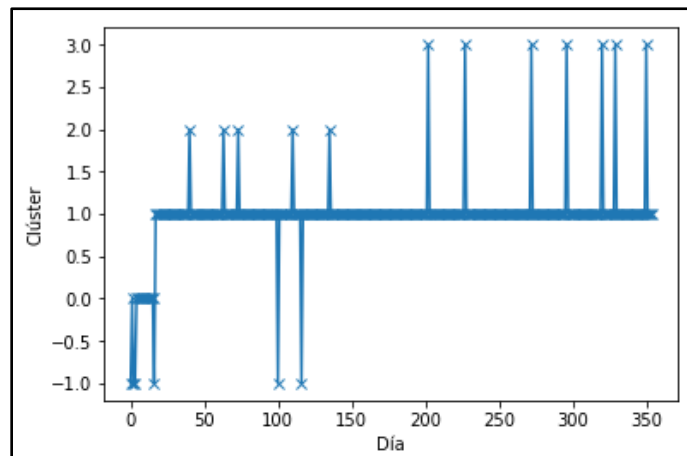
BANCO 20



BANCO 21

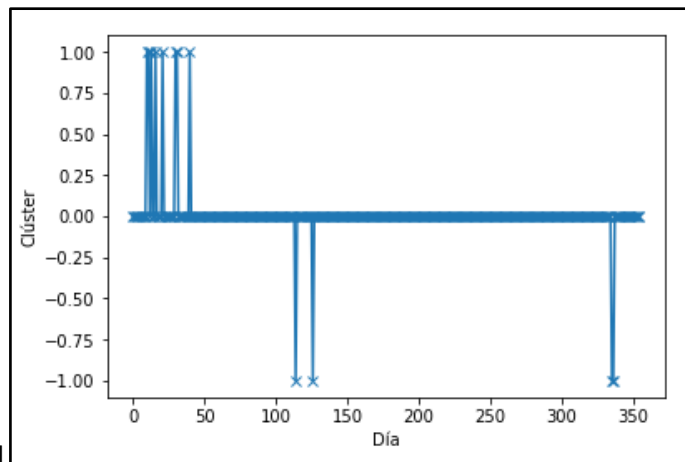


BANCO 22

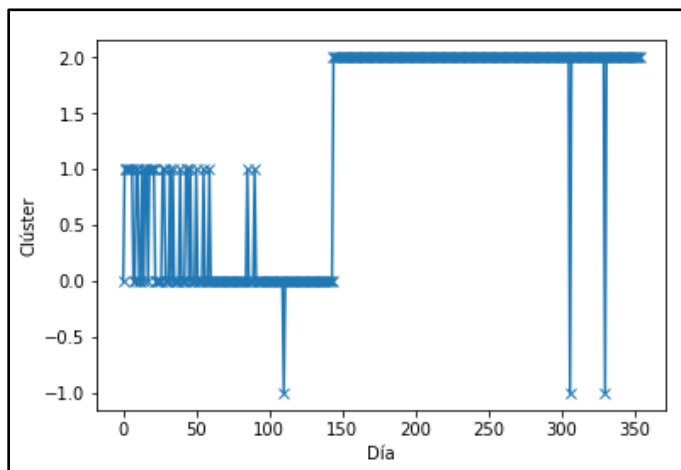


IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

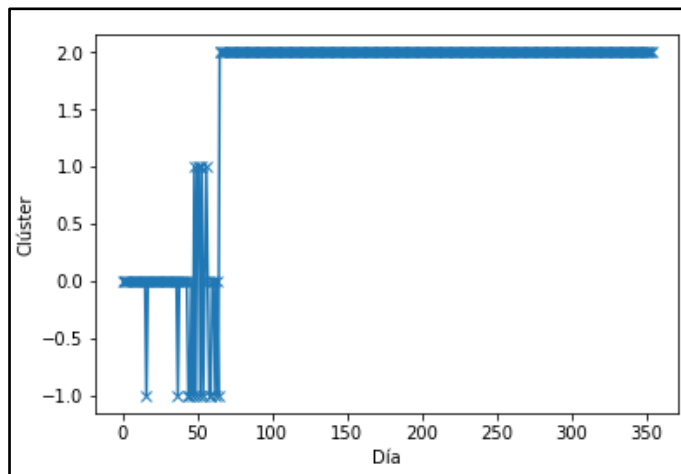
BANCO 23



BANCO 24

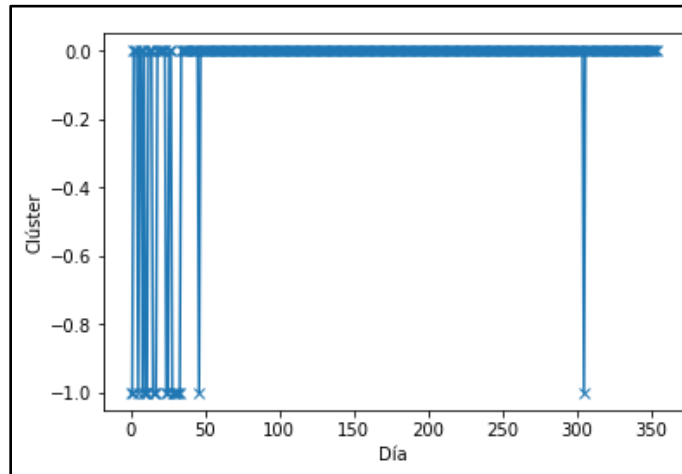


BANCO 25

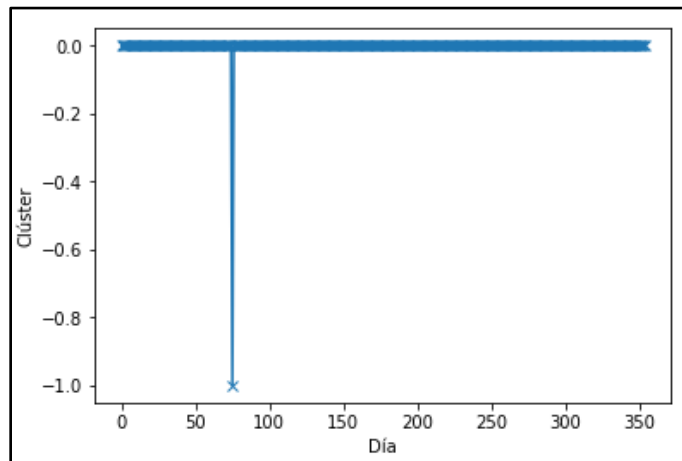


IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

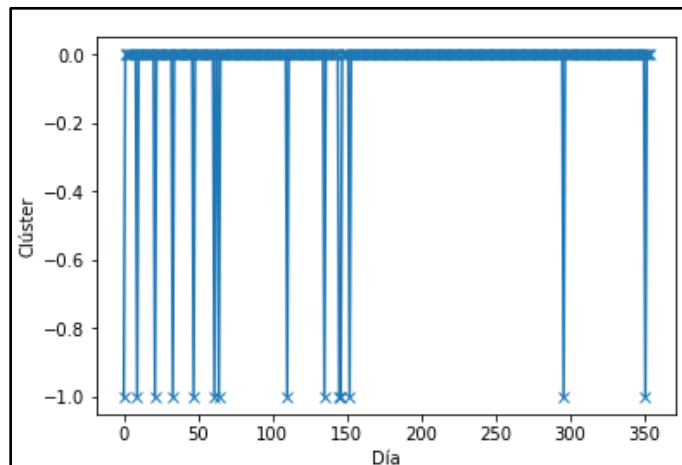
BANCO 26



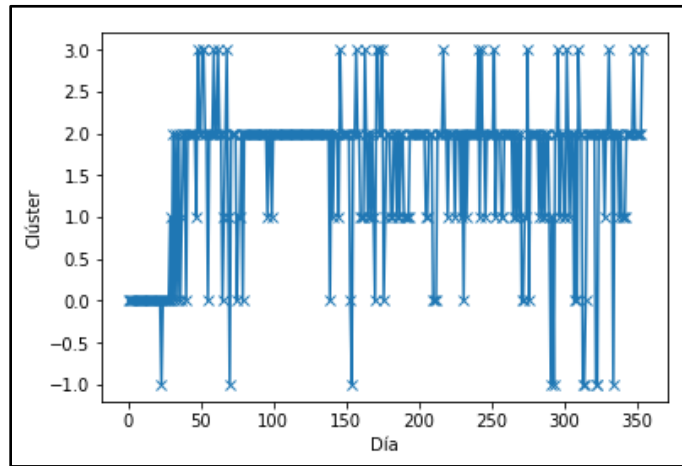
BANCO 27



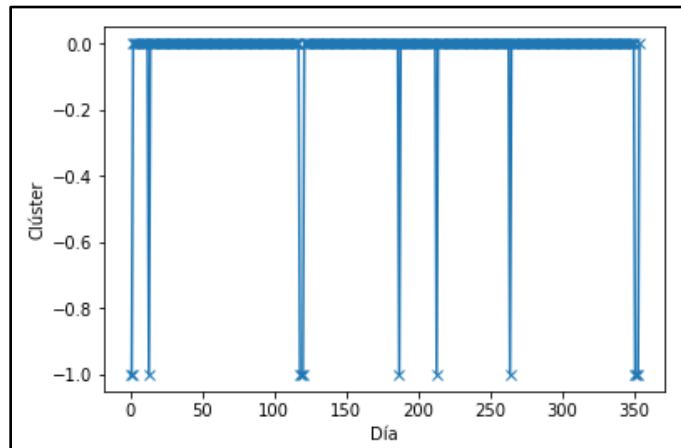
BANCO 28



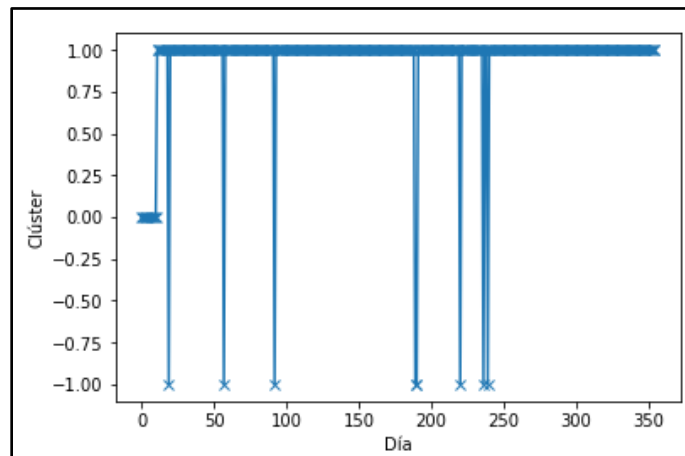
BANCO 29



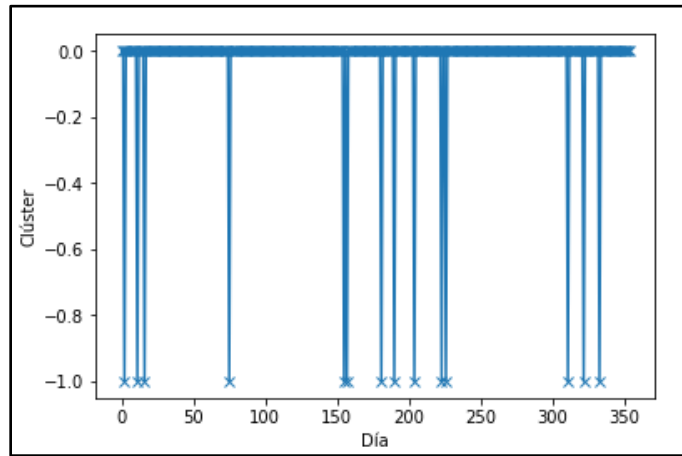
BANCO 30



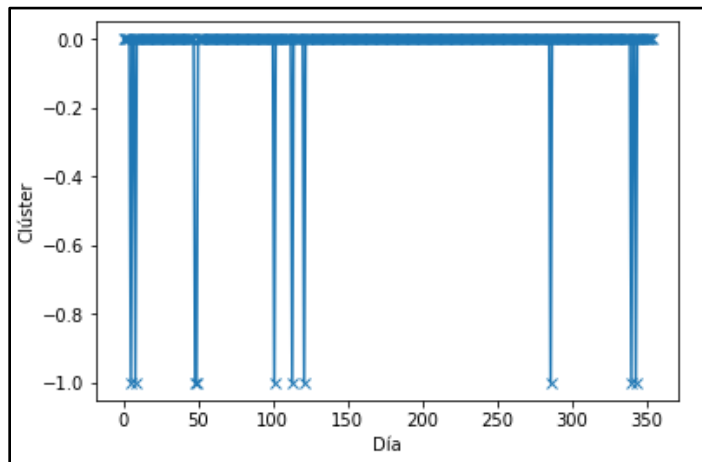
BANCO 33



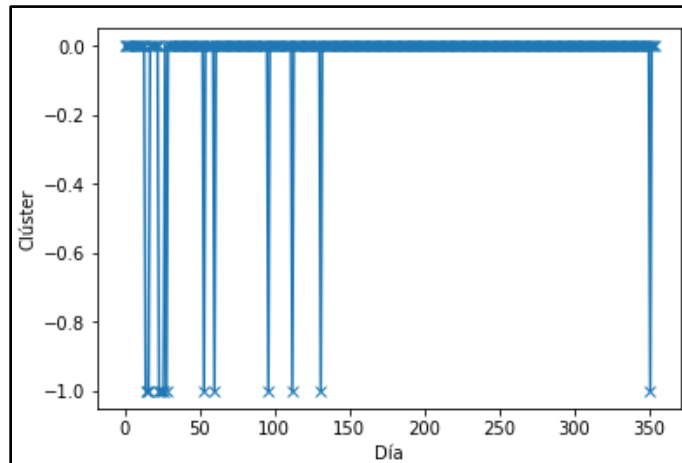
BANCO 34



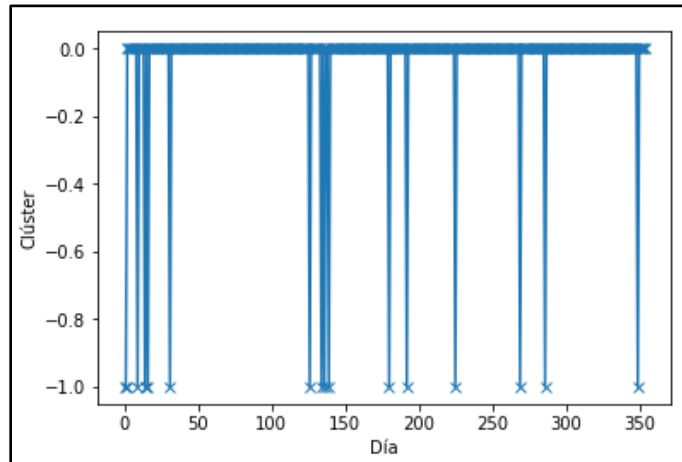
BANCO 35



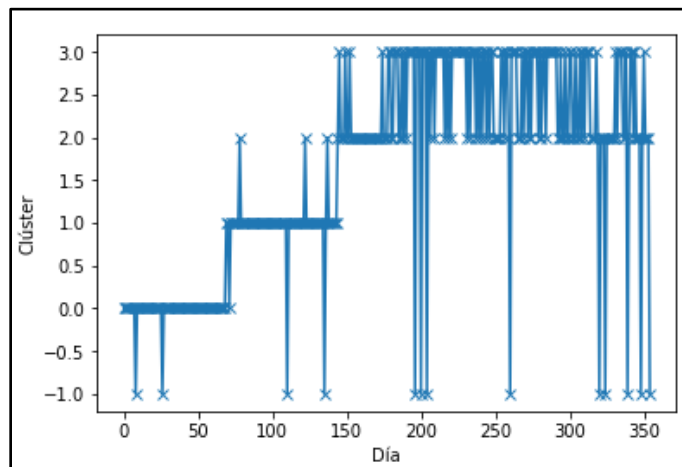
BANCO 36



BANCO 37



BANCO 38



Bibliografía

- Aguilar, J. (Mayo de 2018). *Universidad de los Andes*. Introducción a minería de datos, metodologías y técnicas de minería de datos. Recuperado de: <http://www.ing.ula.ve/~aguilar/actividad-docente/IN/transparencias/clase40.pdf>
- Backlund, H., Hedblom, A., & Neijman, N. (2011). *A Density-Based Spatial Clustering of Application with Noise*. Recuperado de [http://staffwww.itn.liu.se/~aidvi/courses/06/dm/Seminars2011/DBSCAN\(4\).pdf](http://staffwww.itn.liu.se/~aidvi/courses/06/dm/Seminars2011/DBSCAN(4).pdf)
- Banco de México. (2018). Banco de México. Recuperado de <http://www.banxico.org.mx/>
- Barranco, F., Guerrero Galindo, A., & Entrena, M. (2005). *Introducción a Python*. Recuperado de <http://lsi.ugr.es/~pdo/Seminarios/Python2005.pdf>
- Benítez Sánchez, I. J. (2005). Técnicas de agrupamiento para el análisis de datos cuantitativos y cualitativos (Tesis de Grado). Valencia, España.
- Bravo-Benitez, B., Kabadjova, B. A., & Martinez-Jaramillo, S. (2014). *Centrality Measurement of the Mexican Large Value Payments System from the Perspective of Multiplex Networks, [42-44]*. Recuperado de: <https://www.researchgate.net/publication/263889322>
- Dadaser-Çelík, Çelík, M., & Sakir Dokuz, A. (2011). *Revista IEEE. Anomaly Detection in Temperature Data Using DBSCAN Algorithm*. Recuperado
- Ester, M., Krieger, H.-P., Sander, J., & Xu, X. (1996). *Revista AAAI. A density-Based Algorithm for discovering clusters*. Recuperado de <https://www.aaai.org/Papers/KDD/1996/KDD96-037.pdf>
- García Morate, D. (2017). *Manual Weka*. Recuperado de: <https://knowledgesociety.usal.es/sites/default/files/MANUAL%20WEKA.pdf>
- Giménez García, M. (2009). Utilización de Sistemas de Detección de Intrusos como elemento de seguridad perimetral. (Trabajo de grado, Universidad de Almería). Recuperado de http://www.adminso.es/recursos/Proyectos/PFC/PFC_marisa.pdf.
- Gómez, G. M. (06 de Abril de 2016). *CONCACYT*. El poder del aprendizaje automático unido al poder de cálculo de las computadoras actuales. Recuperado de: <http://ccc.inaoep.mx/~pgomez/conferences/PggTSys16.pdf>
- González Bernal, J. A. (2017). *Minería de Datos*. Recuperado de https://ccc.inaoep.mx/~jagonzalez/AI/Sesion13_Data_Mining.pdf
- González, A. (2014). *CleverData*. Recuperado de ¿Que es Machine learning?: <http://cleverdata.io/que-es-machine-learning-big-data/>
- Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques*. Recuperado de: <https://web.engr.illinois.edu/~hanj/bk2/toc.pdf>

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

informática, I. t. (2018). Sistema de detección de intrusos. *Actualidad TIC*. Recuperado de <http://web.iti.upv.es/actualidadtic/2005/02/2005-02-intrusos.pdf>.

Internacionales, B. d. (2001). Comité de sistemas de pago y liquidación (CPSS). Principios Básicos para los sistemas de pago de importancia sistémica. Banco de México. [6-7] Recuperado de <http://www.banxico.org.mx/sistemas-de-pago/material-educativo/intermedio/%7B4029459F-BC12-A46F-EE4B-728F41F2B102%7D.pdf>

Matlab (2018). Mathworks. Recuperado de: <https://la.mathworks.com/products/matlab.html>

MathWoks. (2017). *Matlab*. Recuperado de <https://la.mathworks.com/products/matlab.html>

México, CEMLA. (2013). *Introducción a los sistemas de pago y liquidación*. Recuperado de <http://www.cemla.org/actividades/2013/2013-09-sistemas-de-pagos/2013-09-sistemas-de-pagos-01.pdf>

México, B. d. (2016). *Banxico*. Divulgación del cumplimiento y adopción de los Principios para las Infraestructuras del Mercado Financiero. Recuperado de: <http://www.banxico.org.mx/sistemas-de-pago/material-educativo/intermedio/evaluaciones-conforme-a-las-mejores-practicas-inte/conforme-a-las-mejores-practicas-internacionales/%7BDE853D65-8859-1D34-76EE-0ECF98065A59%7D.pdf>

Morales, E., & Escalante, H. J. (2017). *Clustering*. Recuperado de <https://ccc.inaoep.mx/~emorales/Cursos/NvoAprend/Acetatos/clustering.pdf>

Navas Moreno, F. (2016). *Introducción a la minería de datos con weka: Aplicación a un problema económico*. (Trabajo de grado). Recuperado de: <http://tauja.ujaen.es/bitstream/10953.1/6984/1/TFG%20%20navas%20moreno%2C%20%20Francisco.pdf>

Ning Tan, P., Steinbach, M., & Kumar, V. (2006). *Introduction to Data Mining*. Pearson Addison-Wesley. Recuperado de <https://bayanbox.ir/view/2627688347011855619/IntroductiontoDataMing.pdf>

Pascual González, D. (2010). *Algoritmos de agrupamiento basados en densidad y validación de clusters*. (Tesis Doctoral) Castellón, España.

Santamaría Ruíz, W. (2010). *Modelo de detección de fraude basado en el descubrimiento simbólico de reglas de clasificación extraídas de una red neural*. (Tesis de maestría). UNIVERSIDAD NACIONAL DE COLOMBIA, Bogotá.

Shah, H., Napanda, K., & D'Mello, L. (2015). Revista *International Journal of Computer Sciences and Engineering*. Density Based Clustering Algorithms. Recuperado de http://www.ijcseonline.org/pub_paper/11-IJCSE-01380.pdf

Prado-Romero, A., & Gago-Alonso, A. (2015). *Análisis de los métodos para la detección de anomalías en redes sociales*. La habana, Cuba. Recuperado de https://www.academia.edu/13896724/An%C3%A1lisis_de_los_m%C3%A9todos_para_la_

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN

detecci%C3%B3n_de_anomal%C3%ADas_en_redes_sociales

https://www.academia.edu/19127034/Anomaly_detection_in_temperature_data_using_DBSCAN_algorithm

Python (2018). Python. Recuperado de: <https://www.python.org/>

Pol'la, M., Buccella, A., Cechich, A., & Doldan, M. (2016). *Detección de Patrones de Distribución en Ecología Marina: Un caso de estudio*. Recuperado de: <http://conaiisi.unsl.edu.ar/2013/161-439-1-DR.pdf>

Qayyum, A., Islam, M., & Jamil, M. (2005). Revista IEEE Taxonomy of statistical based Anomaly Detection Techniques for Intrusion Detection. Recuperado de http://iba.edu.pk/IBALibraries/ORC/onlinebooks/int_2005_emer_tech/Files/270.pdf

Romero, I., & Alonso, C. (2017). *Aplicación de técnicas de detección de anomalías a escenarios de ciudades inteligentes*. Recuperado de: <http://ocs.editorial.upv.es/index.php/JITEL/JITEL2017/paper/viewFile/6574/3215>

Spyder. (2017). *Spyder- Documentation*. Recuperado de <https://pythonhosted.org/spyder/>

Spyder (2018). Spyder. Recuperado de: <https://pythonhosted.org/spyder/#>

Uriz Martin, M. X., & Galar Idoate, M. (2015). Aprendizaje de distancias basadas en disimilitudes para el algoritmo de clasificación KNN. (Trabajo de grado, Universidad Politécnica Nacional). Recuperado de <http://academica-e.unavarra.es/xmlui/bitstream/handle/2454/18430/MemoriaTFGMikelUrizMartin-1.pdf?sequence=1>

Wanner, L. (2004). *Introduction to Clustering Techniques*. Recuperado de <http://www.iula.upf.edu/materials/040701wanner.pdf>

Weka (2018). Weka the university of Waikato. Recuperado de: <https://www.cs.waikato.ac.nz/ml/weka/>

Wikiwand. (2018). *Wikiwand*. Análisis de grupos. Recuperado de http://www.wikiwand.com/es/An%C3%A1lisis_de_grupos

IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO FINANCIERO UTILIZANDO EL ALGORITMO DBSCAN
