

MINERÍA DE DATOS

Evaluación Final Unidad I

Nombre: Silvana E. Medina Garrido
Cristian Matus Mallorca

Puntaje total ____ de 42. Nota: ____

Instrucciones: Responda las siguientes preguntas de desarrollo. Sea breve pero certero en sus respuestas. Utilice la terminología correcta para cada expresión.

Respuesta completa	3 puntos
Respuesta incompleta	2 puntos
Sin respuesta o respuesta incorrecta	0 puntos

1.- Explique los siguientes términos relacionados a la Minería de datos (27 puntos [3 puntos c/u]):

A.- KDD (Knowledge Discovery in Databases): Este realiza el proceso extracción de patrones a través de la recopilación de los datos de todas las fuentes y homogenizarlos e integrarlos, esto se ocupa para la toma de decisiones dentro de una organización.

B.- Machine learning: Esta es una forma de inteligencia artificial que permite a un sistema aprender de los datos en lugar de aprender mediante programación explícita, este método de análisis de datos automatiza la construcción de modelos analíticos, se basa en la idea de sistemas que pueden aprender datos, identificar patrones y tomar decisiones con mínima intervención humana.

C.- Análisis Predictivo: Es el proceso de utilizar el análisis de datos para realizar predicciones basadas en estos mismos, este proceso hace uso de los datos analíticos, estadísticas y de aprendizaje automático con el fin de crear un modelo predictivo para predecir eventos futuros, ayudando así a los equipos de sectores diversos como, financieros, sanidad, farmacéuticos.

D.- Análisis Cluster: Es una técnica estadística multivariante que busca agrupar elementos, para lograr la máxima homogeneidad en cada grupo y sus diferencias.

E.- BigData: Es el almacenamiento y procesamiento de grandes cantidades de datos estructurados, semiestructurados y no estructurados, para que luego puedan ser extraídos y organizados una forma que se puedan ocupar para proporcionarnos información valiosa.

F.- ETL: Extract, Transform and Load. Es un proceso de gestión de datos de tres pasos que extrae datos no estructurados de distintas fuentes, transformadas en un formato que satisfaga los requisitos operativos y analíticos del negocio.

G.- Instancia, Atributos:

Instancias: elemento atómico de información proveniente de un set de data. Conocido también como récords, prototipos o ejemplos.

Atributos: medidas de aspectos de una instancia. Conocido también como características o variables. Cada instancia está compuesta de un cierto número de atributos.

H.- Herramientas de Limpieza: Son herramientas que utilizan un conjunto de técnicas para explorar grandes cantidades de datos

I.- Representación del conocimiento: Representar el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y, en última instancia, comprensibles a partir de datos.

2.- Mencione 3 campos en los que es aplicable la Minería de Datos y de un ejemplo para cada uno (3 puntos).

Respuesta: La minería de datos puede ser aplicado en diferentes campos y/o áreas como pueden ser los siguientes campos:

Marketing: En el campo del marketing se puede utilizar de diferentes formas mediante los datos adquiridos de las diferentes campañas publicitarias, el resultado de estas, la cantidad de clientes que llegan a consumir un producto determinado así demostrando la efectividad de un cierto tipo de publicidad contra un cierto sector de clientes. Esto permite generar grandes ahorros en lo que refiere a campañas publicitarias, reduciendo el factor de “prueba y error” que se podría presentar en muchos casos.

Medicina: En cuanto al campo de la medicina, podemos observar que la minería de datos se puede representar de diferentes formas, entre una de ellas se encuentra en la detección de patrones en los casos de desarrollo de enfermedades de diversas gravedades, lo cual permite y facilita la elección de diversos tratamientos para prevenir o tratar estas enfermedades.

Entretenimiento: Por último, en el campo del entretenimiento es uno de los campos donde se puede sacar múltiples beneficios a la minería de datos, los cuales pueden consistir en mejorar la eficiencia de los recursos destinados a diferentes proyectos, la capacidad de consumo que puede obtener los diferentes grupos de clientes objetivos, etc.

3.- Indique y explique brevemente cuál es el proceso (fases) de la Minería de Datos (3 puntos).

Respuesta: La minería de datos, hoy en día se generan grandes cantidades de datos de todo tipo debido a que cada vez más los seres humanos tienen una mayor acceso a los medios digitales en donde no solo se consulta información sino también se comparte y prácticamente todas las organizaciones tienen sistemas informáticos donde van almacenando su información es por ello que la importancia de la minería de datos amigos se basa en descubrir patrones tendencias e historias que se ocultan dentro de cualquier base de datos y que esos descubrimientos se puedan convertir en información relevante.

Las fases:

-filtrado de datos: mediante este proceso se filtran los datos de forma que se eliminan los valores incorrectos, no validos o desconocidos, según las necesidades y el algoritmo a usar.

-selección de variables: tenemos dos métodos de selección de características aquellas basados en la elección de los mejores atributos del problema y aquellos que buscan variables independientes mediante test de sensibilidad, algoritmos de distancia o heurísticos.

-extracción de conocimientos: mediante una técnica de minería de datos, se obtiene un modelo de conocimiento, que nos representa patrones de comportamiento observados en los valores de las variables del problema o relaciones de asociación entre dichas variables.

-interpretación y evaluación: ya una vez obtenido el modelo se debe proceder a su validación, comprobando que las conclusiones que nos arroja son válidas y suficientemente satisfactorias.

4.- Responda a las preguntas presentadas a continuación del siguiente caso (9 puntos [3 puntos c/u])

Caso: Apertura de una sucursal bancaria - La apertura de un punto transaccional requiere la combinación de varios elementos que se configuran en el territorio; dichos elementos son capaces de condicionar zonas de oportunidad o de desventaja. Dentro de los elementos están en primer lugar la presencia de microempresarios, en donde se asume que la decisión de ir a un sitio "x", está determinado tanto por la distancia como por el tiempo, por lo tanto, los hábitos de transaccionalidad bancaria tienen un componente geográfico importante, lo que a priori determina la posición óptima de un punto de Sucursal Bancaria. Por otro lado, están los antecedentes sociodemográficos, tales como la segmentación socioeconómica y demografía, que permiten un marketing selectivo, dividiendo a los potenciales clientes en grupos a partir de la aplicación de criterios de discriminación tales como cantidad de uso y nivel transaccional, para luego desarrollar planes de captación orientados a cada uno de los grupos resultantes. Dada la necesidad de evaluar la posibilidad de apertura de una sucursal bancaria, entonces, como ingeniero informático que realiza labor de minería de datos, se requiere saber:

**A.- ¿Es este un tipo de problema que podría ser resuelto con minería de datos?
¿Por qué?**

Si, ya que la minería de datos busca explorar bases de datos que son cada vez más, analizando distintos parámetros para interpretar los resultados o de otra forma, recolectando datos para luego interpretarlos y tener una fuente de información con relación a un tema en concreto.

B.- Plantee mínimo 10 atributos (nivel socioeconómico, cantidad de personas, etc.) y argumente por qué sería importante poseer estos tipos o conjunto de datos, es decir, qué información le permitiría obtener.

- 1) Clase social
- 2) Ingreso promedio
- 3) Empleo
- 4) Estatus social
- 5) Estratificación social
- 6) cantidad de empresas en el sector
- 7) existencia de otros bancos
- 8) cesantía
- 9) edades
- 10) delincuencia

Creo que todos estos datos darían una visión completa de aspectos importantes de la población, ya que por un lado vemos la existencia de competencia, el riesgo y la oportunidad que tenemos al establecer nuestra sucursal en el área.

Tabla de puntajes:

Puntaje	Nota
0.0	1.0
1.0	1.1
2.0	1.2
3.0	1.4
4.0	1.5
5.0	1.6
6.0	1.7
7.0	1.8
8.0	2.0
9.0	2.1
10.0	2.2
11.0	2.3
12.0	2.4
13.0	2.5
14.0	2.7

Puntaje	Nota
15.0	2.8
16.0	2.9
17.0	3.0
18.0	3.1
19.0	3.3
20.0	3.4
21.0	3.5
22.0	3.6
23.0	3.7
24.0	3.9
25.0	4.0
26.0	4.1
27.0	4.3
28.0	4.5
29.0	4.7

Puntaje	Nota
30.0	4.9
31.0	5.0
32.0	5.2
33.0	5.4
34.0	5.6
35.0	5.8
36.0	5.9
37.0	6.1
38.0	6.3
39.0	6.5
40.0	6.6
41.0	6.8
42.0	7.0