

UNIVERSIDAD PRIVADA FRANZ TAMAYO

**FACULTAD DE INGENIERÍA
CARRERA DE INGENIERÍA DE SISTEMAS**



Documento de práctica primer parcial minería de datos

Estudiantes: Erick Rodrigo Aranibar Arias
Einar David Villarroel Vargas
René Rocha Antezana

Docente: Miguel Ángel Paco

Materia: Minería de datos

Cochabamba-Bolivia
2019

Introducción.

El presente documento trata sobre la realización de la práctica del primer parcial de la materia de minería de datos. Esta práctica consistió en la realización de un software, en código python, el cual permita la recopilación de datos respecto a qué temas suelen hacer felices, tristes y preocupadas a las personas tanto optimistas como pesimistas.

Desarrollo.

Para la realización de la práctica fue necesaria la recolección de datos por medio de una encuesta. Esta encuesta se realizó por medio de un software web con el motor Node js y constó de las siguientes preguntas:

¿Eres Una Persona Feliz O Triste?

Menciona 1 o 2 actividades que te hagan feliz

Escribe aqui

Menciona 1 o 2 actividades que te hagan sentir triste

Escribe aqui

Menciona 1 o 2 actividades que te hagan sentir preocupado

Escribe aqui

Te consideras una persona Optimista o Pesimista

☐ Optimista

☒ Pesimista

Enviar Datos

La encuesta aún puede ser consultada por medio del siguiente enlace:

<https://mineriaoptimismo.herokuapp.com/>

Las preguntas fueron recogidas por medio de documentos txt disponibles junto con el código de python en el siguiente enlace:

<https://github.com/friki57/encuestasMineria>

Una vez fueron recogidos los datos se proceso de categorización en el cual se agruparon las respuestas de la encuesta en las siguientes categorías:

familia, arte, ocio, social, estudios, deporte y ejercicio, animales, naturaleza, salud, trabajo y no sabe o no responde.

Las categorías siguieron una lógica en la cual cada una está asociada con un conjunto de palabras claves, de modo que, si alguna de estas palabras aparecía en la respuesta entonces esta respuesta entraba en la categoría correspondiente.

Las palabras claves de cada categoría son:

familia =

['papas','padres','madre','padre','hermanita','familia','familias','papá','mamá','herman','primo','prima','abuel','casa']

arte =

['musica','música','museo','bailar','baile','cantar','canto','canciones','guitarra','piano']

ocio =

['comer','dormir','videojuegos','television','televisión','tv','internet','películas','peliculas','celular','auto','free fire','fumar','viaje','viajar','computador','leer']

social =

['amigo','amiga','compañeros','fiesta','amigos','solo','sola','ayudar','personas','cañar','indiferen','discut','discusión','mentir','soledad','novi']

estudios =

['notas','universidad','colegio','clase','materia','exámenes','examen','exámen','carrera','estud','bachiller','tarea','exponer','profesion','profecion','profesión','instituto','hito','laboratorio','cole']

deporte y ejercicio =

['deporte','ejercicio','fútbol','futbol','gym','gimnasio','voleibol','basket','basquet','camina','crossfit']

animales = ['animal','perr','gato','gata','toro']

no sabe o no responde = ['nada','ningun','ningún']

naturaleza =

['planta','viento','chiquitania','ambiente','planeta','amanecer','anochece','atardecer','contamin','chaqueo','estrella']

salud =

['enfermedad','salud','cáncer','enfermar','hospital']

trabajo =

['lavar','trabaj','laburar','empleo','dinero','económico','econom','pobre','sueldo','deudas']

El código de python finalmente tras realizar la categorización de las respuestas llega a generar tres tablas de machine learning. Estas tablas son correspondientes a cada pregunta y muestran las distintas categorías que coinciden como sus atributos y a la variable objetivo como si la persona es optimista o pesimista.

Las tablas también se pueden encontrar en el mismo repositorio de github que el código, en forma de 3 archivos .xlsx, de todos modos las tablas son las siguientes:

[illegible]

[illegible]

[illegible]

Conclusiones.

Al final la encuesta fue realizada por un total de 102 personas, de las cuales se obtuvieron respuestas bastante variadas, las cuales ayudaron a formar la lista de palabras clave para la categorización. La categorización resultó bastante efectiva, a tal punto que cerca de 90 de las 102 encontraron al menos una coincidencia, dejando un 10% de respuestas que no pudieron ser categorizando correctamente.

De estas respuestas se pueden sacar conclusiones como definir ¿Qué cosas son las que hace feliz a una persona optimista? Y ¿Qué cosas son las que hacen triste a las personas pesimistas?