**Informe Ejecutivo**

**Evaluación 3**

**Integrantes:**

Oscar Espinoza

William Menares

Patricio Gonzalez

**Docente:**

Carlos Correa

***Introducción***  
**Fase 1: Entendimiento del negocio:**

**Contexto:**

Somos inversionistas de una empresa minera de oro que busca realizar un estudio sobre el clima en Australia, con el fin de averiguar en qué sectores llueve más y en qué sectores llueve menos, esto para encontrar un lugar adecuado donde instalar una nueva planta minera de extracción y procesamiento del mineral, ya que tanto los sectores de abundante lluvia, así como también los sectores más áridos pueden tener un impacto negativo en la extracción de minerales. Estos problemas pueden ser los siguientes:

**Impacto en las operaciones de extracción:**

* Paralización de actividades por inundaciones.
* Riesgos de deslizamiento.
* Contaminación de minerales.
* Escasez de agua en para extracción de minerales que la requieran zonas áridas.
* Condiciones extremas de calor en zonas áridas.

**Impacto en la infraestructura:**

* Erosión de caminos debido a mucha lluvia.
* Estabilidad de Estructuras en zonas de mucha lluvia.
* Desgaste de infraestructura por temperaturas extremas.
* Corrosión por polvo en zonas áridas
* Abastecimiento de agua en zonas áridas.

**Impacto en el procesamiento de minerales:**

* Dilución de soluciones de lixiviación por exceso de agua con barro debido a lluvias fuertes.
* Corrosión de equipos por lluvias constantes y humedad.
* **Consumo de agua:** Los procesos de beneficio de minerales a menudo requieren grandes cantidades de agua, que pueden ser difíciles de obtener en áreas áridas.

**Impacto en la logística y transporte:**

* Bloqueo de Rutas de Acceso.
* **Impacto en vehículos:** El polvo y las temperaturas extremas pueden acelerar el desgaste de los vehículos.
* **Problemas con el transporte:** problemas con el transporte al interior del sitio minero así como problemas con costos más altos de transportes o retrasos en la entrega.

**Objetivo del proyecto:**

* El objetivo principal del proyecto es identificar las regiones en Australia con las condiciones climáticas más adecuadas para la instalación de una planta minera de oro. Esto implica encontrar un equilibrio climático que minimice los impactos negativos en las operaciones.

**Necesidades de la empresa:**

**Para alcanzar el objetivo de nuestro proyecto, necesitamos:**

* **Datos climáticos detallados:** Información precisa y actualizada sobre las precipitaciones, temperaturas y otros factores climáticos en diferentes regiones de Australia.
* **Análisis de impacto climático:** Evaluación de cómo las diferentes condiciones climáticas pueden afectar las operaciones mineras, infraestructura, procesamiento de minerales y logística.
* **Proyecciones climáticas:** Modelos y predicciones a futuro que permitan entender cómo podrían cambiar las condiciones climáticas y su impacto potencial en las operaciones mineras.
* **Recomendaciones estratégicas:** Sugerencias basadas en el análisis de datos que identifiquen las mejores regiones para la instalación de la planta minera.

***Desarrollo***

**Fase 2: Entendimiento de los Datos**

Comprensión inicial de los datos disponibles para identificar características clave de los datos, detectar posibles problemas y orientar el análisis posterior.

**Actividades:**

* **Exploración inicial:** Se realizó una inspección de las primeras filas del dataset para entender la estructura de los datos.
* **Resumen estadístico:** Se generaron estadísticas descriptivas para todas las variables, incluyendo media, mediana, desvío estándar, etc.
* **Detección de valores faltantes:** Se identificaron las variables con valores nulos y se analizó la cantidad y el porcentaje de estos.
* **Análisis de correlación:** Se creó un mapa de calor para visualizar las correlaciones entre variables.
* **Visualización de distribuciones:** Se usaron histogramas y gráficos de densidad para analizar la distribución de las variables.
* **Visualización de mapa:** Se utilizó una API para obtener coordenadas de las ciudades de australia

**Fase 3: Preparación de los Datos**

Procesamiento de los datos para dejarlos listos para el modelado para asegurar que los datos sean adecuados para el análisis, eliminando o transformando valores que podrían afectar el rendimiento de los modelos.

**Actividades:**

* **Limpieza de datos:** Se eliminan duplicados y se manejaron valores extremos (outliers).
* **Imputación de valores faltantes:** Utilizamos diferentes métodos (media, mediana, moda) para rellenar los valores nulos.
* **Codificación de variables categóricas:** Aplicamos la técnica One-Hot Encoding para convertir variables categóricas en numéricas.
* **Escalado de variables:** Se normalizaron o estandarizaron las variables para que todas tengan una escala comparable.
* **Balanceo de clases:** Usamos SMOTE para balancear la distribución de las clases en problemas de clasificación.
* **Segmentación de datos:** Separamos los datos separando en las ciudades de la zona del país que más nos interesan para el negocio para luego modelar con esos datos.

**Fase 4: Modelado**

**Modelos de Regresión**

Creación y evaluación de modelos de regresión para predecir variables continuas para realizar predicciones cuantitativas y obtener insights sobre las variables que más influyen en la variable objetivo.

**Actividades:**

* **Selección de variables:** Se identificaron las variables más relevantes mediante análisis de correlación y técnicas de selección de características.
* **Creación de modelos:** Realizamos modelos de regresión lineal múltiple , KNNRegressor y Random forest.
* **Evaluación de modelos:** Se evaluaron los modelos utilizando métricas como el RMSE , el MAE y Coeficiente de Determinación
* **Selección de Modelo:** El mejor modelo de regresión es el KNN Regressor debido a que tiene un menor error a comparación de los demás, y tiene un mayor porcentaje de Precisión.

**Modelos de Clasificación**

Creación y evaluación de modelos de clasificación para predecir categorías para clasificar correctamente las observaciones en categorías, mejorando la toma de decisiones basada en datos categóricos.

**Actividades:**

* **Selección de variables:** Seleccionamos las variables más influyentes para el target de clasificación.
* **Creación de modelos:** Aplicamos modelos Regresión Logística, Random Forest, Ada Boost y Naive Bayes.
* **Evaluación de modelos:** Se evaluaron los modelos utilizando Curva de Roc y Validación Cruzada.
* **Selección de Modelo:** El mejor modelo de clasificación es el Ada Boost debido a que tienen un precisión y recall más equilibrado, esto quiere decir que Ada Boost tiene un rendimiento más uniforme y balanceado, detectando mejor los positivos y negativos.

**Clustering**

Aplicación de técnicas de clustering para agrupar datos sin etiquetas para descubrir patrones ocultos y segmentar los datos en grupos significativos.

**Actividades:**

* **Selección de características:** Se seleccionaron las variables más relevantes para el análisis de clustering.
* **Creación de modelos de clustering:** Se implementaron modelos KMeans, GaussianMixture, AffinityPropagation y AgglomerativeClustering.
* **Evaluación de clustering:** Utilizamos las métricas puntaje de silueta e índice de Davies-Bouldin para evaluar la calidad de los clusters.
* **Selección Modelo:** El mejor modelo de clustering es el K Means debido a que presentó mejores métricas de silhouette, Davis Bouldin y presentó mejor segmentación y separación de sus clusters centrales.

***Conclusión***

En conclusión podemos decir que aunque tuvimos unos desempeños medianos, faltan datos en la zona elegida , ya que nosotros como minera de carbón elegimos la zona Este de australia donde se presentan las mayores concentraciones de carbón y los datos que nosotros teníamos de esa zona solo eran de 3 ciudades las cuales eran 8700 datos aproximadamente, entonces no podemos conseguir un desempeño alto en el modelado.

Sin embargo pudimos analizar diferentes factores de esa zona, y vimos que no existe tanta concentración de las lluvias, tampoco es un ambiente seco, por lo tanto consideramos que la mejor ciudad debido a las condiciones climatológicas fue Moree ya que de las 3 ciudades es la que menos tiene concentraciones de lluvias