**Grupo 3. Avances del proyecto, documentación de los modelos entrenados**

**YOSELIN NIETO GIL, FELIPE DIEGO FELIPE DIEGO LOBATO DA SILVA, NICOLAS ALEJANDRO YEPES JOVEN y ANA MARIA SOTO OROZCO**

Con la intensión de determinar el mejor modelo para resolver las necesidades de la empresa Olist Store, de definió desarrollar modelo que predigan el precio de productos de acuerdo a la temporada en la que se compran, la ubicación del cliente, la categoría del producto y sus características físicas. Sobre estos modelos se aplicó validación cruzada y grindsearch para encontrar los parámetros óptimos en cada modelo y se comparó el desempeño en distintas métricas como RMSE, MAE y R2:

1. **Modelo con XGBRegressor:**
   * **Descripción**: Se implementa un pipeline que incluye preprocesamiento y el modelo XGBRegressor. Se realiza una búsqueda de hiperparámetros con GridSearchCV.
   * **Parámetros óptimos**: {'colsample\_bytree': 0.8, 'gamma': 0.1, 'learning\_rate': 0.1, 'max\_depth': 9, 'n\_estimators': 300, 'reg\_alpha': 0.01, 'subsample': 0.8}
   * **Métricas**:
     + **RMSE**: 132.67
     + **MAE**: 54.75
     + **R²**: 0.5368
     + **RMSE (validación cruzada)**: 127.56
2. **Modelo con RandomForestRegressor:**
   * **Descripción**: Similar al anterior, este pipeline utiliza RandomForestRegressor y RandomizedSearchCV para ajustar hiperparámetros.
   * **Parámetros óptimos**: {'n\_estimators': 100, 'min\_samples\_split': 5, 'min\_samples\_leaf': 1, 'max\_features': 'log2', 'max\_depth': None, 'bootstrap': False}
   * **Métricas**:
     + **RMSE**: 135.53
     + **MAE**: 43.09
     + **R²**: 0.5167
     + **RMSE (validación cruzada)**: 132.04
3. **Modelo con LinearRegression:**
   * **Descripción**: Modelo lineal básico, sin ajuste de hiperparámetros. Incluye preprocesamiento mediante un pipeline.
   * **Métricas**:
     + **RMSE**: 118.01
     + **MAE**: 25.55
     + **R²**: 0.6336
     + **RMSE (validación cruzada)**: 114.60
4. **Modelo con ElasticNet:**
   * **Descripción**: Modelo de regresión penalizada (ElasticNet) que también se entrena en pipeline.
   * **Métricas**:
     + **RMSE**: 183.53
     + **MAE**: 81.10
     + **R²**: 0.1137
     + **RMSE (validación cruzada)**: 171.00
5. **Modelo con KNeighborsRegressor:**
   * **Descripción**: Modelo KNeighborsRegressor con ajuste de hiperparámetros mediante GridSearchCV.
   * **Parámetros óptimos**: {'n\_neighbors': 5, 'p': 1, 'weights': 'distance'}
   * **Métricas**:
     + **RMSE**: 142.54
     + **MAE**: 39.61
     + **R²**: 0.4654
     + **RMSE (validación cruzada)**: 137.98

**Modelo con Mejor Desempeño**

Basado en el RMSE y el R², el modelo de **regresión lineal (LinearRegression) tiene el mejor desempeño** con un **RMSE de 118.01** y un **R² de 0.6336**. Este modelo proporciona las predicciones más precisas en este conjunto de pruebas y validación cruzada.