

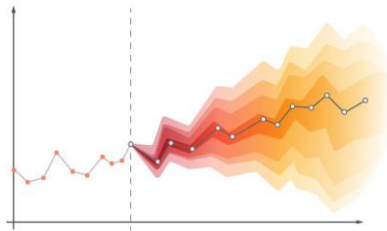
Evaluación Técnica para Científico de Datos: Iniciativa "Driver Score"



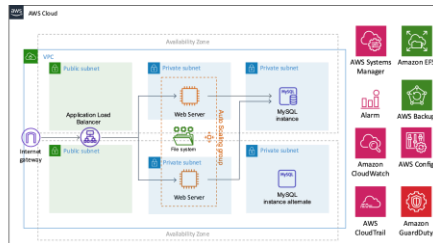
CONTENIDO



Análisis Exploratorio e Ingeniería de Características

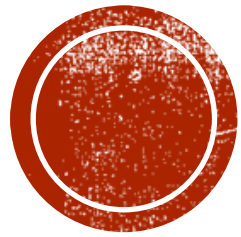


Desarrollo del Modelo Predictivo ("Driver Score")

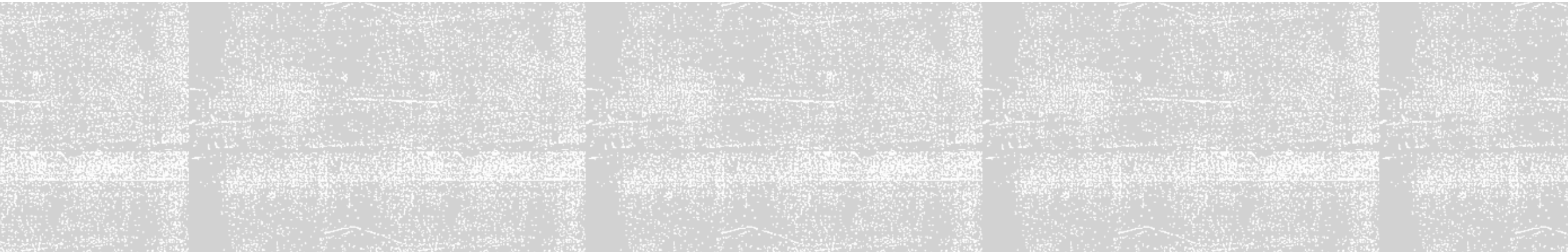


Diseño Arquitectónico para la Industrialización





ANÁLISIS EXPLORATORIO E INGENIERÍA DE CARACTERÍSTICAS



Preprocesamiento y limpieza

Base de datos de viajes del mes de mayo para una empresa real.

163.798



- Nombre
- Fecha
- Ubicación
- Tipo de vehículo
- Información de viaje (Hora de inicio, Hora fin, Total tiempo detenido, Total tiempo manejo, Distancia de viaje, otros)

Limpieza de datos



- Ajuste variables tiempo
- Visualización distribución variables viaje

Creación nuevas variables



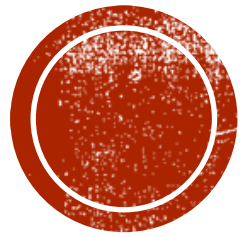
Velocidad esperada según distancia/tiempo, Diferencia entre velocidad máxima y promedio, aceleración media, Magnitud de eventos bruscos, Conteo total de eventos bruscos, Proporción de exceso de velocidad (>80), Viajes con mucha variación entre velocidades, Volatilidad de velocidad y otras

Variable Objetivo



Viajes que cumplen con Eventos bruscos, mucha variación entre velocidades y exceso de velocidad

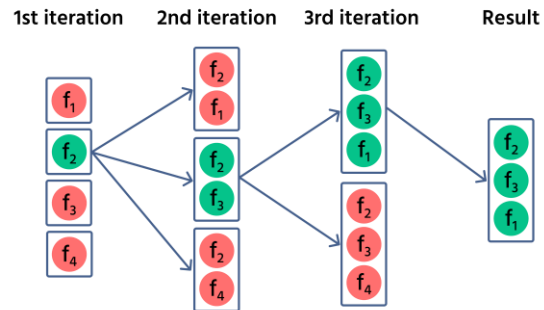




DESARROLLO DEL MODELO PREDICTIVO ("DRIVER SCORE")

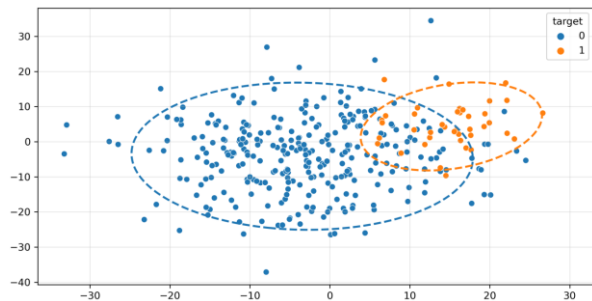


Selección de variables por Backward stepwise y correlaciones



'accel_proxy'
'speed_volatility'
'smoothness_index'

Desbalanceo de Clases



Train - 75%
Test - 25%

Modelo

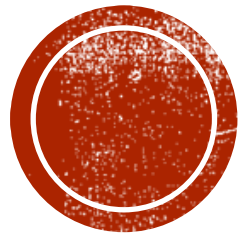
```
automl = AutoML(  
    mode="Perform",  
    ml_task="binary_classification",  
    algorithms=["Random Forest",  
               "Xgboost", "CatBoost", "Neural  
               Network"],  
    eval_metric="f1",  
    results_path="Filter_Model_8",  
    total_time_limit=600,  
    random_state=10,  
    validation_strategy={  
        "validation_type": "kfold",  
        "k_folds": 5,  
        "shuffle": True  
    }  
)
```

MODELAJE

Resultados

| | score | threshold |
|-----------|-----------|-------------|
| logloss | 0.0541147 | nan |
| auc | 0.998352 | nan |
| f1 | 0.977683 | 0.489849 |
| accuracy | 0.980291 | 0.489849 |
| precision | 1 | 0.999859 |
| recall | 1 | 5.97932e-17 |
| mcc | 0.960052 | 0.489849 |





DISEÑO ARQUITECTÓNICO PARA LA INDUSTRIALIZACIÓN



