Project Plan

BAH Machine Learning

Team 3 – Analytics Avengers:

Ahmed Sodeinde

Erick Orellana Morales

Rami Tello

Objective

Develop a model that can be used to predict if a car insurance claim will be filed based on policy features related primarily to the vehicle.

Plan

Step 1: Gather the data - Complete

- Review Kaggle
- Get instructor's approval

Step 2: Setup AWS environment - Complete

- Create users and assign user roles
- Setup S3 buckets with needed permissions
- Setup a SageMaker Notebook and the files for working
- Importing data into SageMaker

Step 3: Data Exploration and Cleanup - Complete

- Understand the data's structure and components
- Descriptive statistics and visualizations
- Correlation analysis
- Initial feature selection
- Handling an unbalanced dataset

Step 4: Create a baseline model - Complete

- Build a dummy model
- Test the model
- Get evaluation statistics for the model

Step 5: Create a traditional linear regression model with feature selection - Complete

- Normalize, encode, and prepare data for modeling
- Create test/train split
- Apply technique to balance dataset
- Apply feature selection
- Build the model
- Evaluate the model

Step 6: Manually create an XGBoost model - Complete

- Normalize, encode, and prepare data for modeling
- Create test/train split
- Apply technique to balance dataset
- Build the model
- Evaluate the model
- Apply hyperparameter tuning
- Build model with optimized hyperparameters
- Evaluate the model

Step 7: Create a model using SageMaker AutoML (Version 4) – Complete

- Setup a domain and run Studio
- Upload dataset into S3 and setup output bucket
- Setup the experiment
- Run the experiment in "Auto" training method
- Evaluate results of the experiment
- Run the experiment in "Hyperparameter Optimization" training method
- Evaluate the experiment

Step 8: Compare and select the most appropriate model – Complete

- Review the evaluation metrics for the models (special attention to F1 score)
- Choose the best model based on evaluation metrics and usability of the solution

Step 9: Deploy the selected model – Complete

- Deploy the selected model (AutoML)
- Pass new data to the endpoint
- Evaluate the results

Step 10: Complete documentation and demo the solution – Complete

- Complete the presentation PowerPoint
- Cleanup code and commit to GitHub
- Complete retrospective