

Case study - Poshmark – Predicting shipping score using AWS Sagemaker



Problem Statement

Poshmark is a leading social commerce platform for the next generation of retailers and shoppers. Goods ordered by customer on the platform are delivered by registered seller. Poshmark wanted to predict the probability of a seller to ship the item to shopper.

Proposed Solution

According to the defined Problem Statement there are two classes in the data “shipped” and “not shipped”. Logistic Regression throws predictions as output which can be consumed in different ways.

Since this data set was highly unbalanced (Shipped-94%) and (Not Shipped 6%), XGBoost was used with data up sampled with SMOTE and without SMOTE.

As the datasets were ready, it was important for Poshmark to predict the cases where goods were “not shipped” accurately which amounted to 6%, hence during the course of the project multiple models were used.

How AWS Services were used to solve this problem

Data was moved from Amazon Redshift to AWS S3. AWS Glue was used to run data cleaning ETL jobs and get data ready in the output format as required so that this can be used by ML models to consume.

Amazon Sagemaker was used to build, train and deploy, machine learning models. Sagemaker Studio gives you complete access, control, and visibility into each step required to build, train, and deploy models.

Customer was able to quickly upload data, create new notebooks, train and tune models, move back and forth between steps to adjust experiments, compare results, and deploy models to production all in one place, making you much more productive. Outcome was sent through API gateway for their apps to consume.

Start Date – October 1, 2018

End Date – January 10, 2019



POSHMARK

About Poshmark

Poshmark is a leading social commerce platform for the next generation of retailers and shoppers.

They're transforming commerce by combining the power of people and technology, making shopping and selling simple, social, and fun.



Outcomes & results

Using Amazon Sagemaker, Poshmark was easily able to create and train a number of specific, targeted models, rather than building a single algorithm.

Poshmark was able to experiment with about 20 variants of algorithm to the defined binary classification problem. This was quickly made possible by Amazon Sagemaker.

After 20 experiments, Poshmark was able to achieve the precision of 98% and Recall of 79% (Predicted as shipment not done / Total actual shipment not done) on these models.

Lessons Learned

It's important to treat each machine learning problem uniquely. Each customer has different data sets, business processes and offerings for their customers.

It's important not to introduce human bias and perform feature engineering objectively by determining features that can impact the outcome.

Ensure that data cleaning, identification of missing data, feature correlation, identification of outliers, and balancing the data is performed effectively.

About Mactores Cognition Inc

Mactores quickly solves core business problems and drives disruptive change by applying the latest automation technologies in Data Analytics, AI/ML and DevOps. We design, deploy, integrate and manage rapid migration and transformation solutions to accelerate enterprise data platform migrations using automation developed over dozens of successful use cases.



MACTORES