Prakhar Goyal goyalpr@mail.uc.edu (513)-807-2655 linkedin.com/in/prakhar3949 prakhar3949.github.io

EDUCATION

University of Cincinnati, Carl H. Lindner College of Business, Cincinnati, Ohio — Expected August 2020
Master of Science in Business Analytics — GPA: 3.96/4
The LNM Institute of Information Technology, Jaipur, India — May 2014
Bachelor of Technology, Electronics and Communication

PROJECTS

Mercari Price Challenge [Tech stack: Python, LGBM, scikit-learn]
- Developed price recommendation model for the sellers by using Gradient boosting(RMSE 0.411,Top 100 Kaggle Public leaderboard).
- Performed hyper parameter tuning using regressor techniques in LGBM library.
- Performed Topic modeling using LDA after pre-processing and converting to sparse matrices (using tf-idf).

Bitcoin Price Prediction System [Tech stack: R, RShiny]
- Developed a Seasonal ARIMA model to forecast next quarter's prices of bitcoin.
- Performed EDA to get insights and differencing orders of the model.
- Created an application on RShiny.

Movie Recommendation System [Tech stack: Python, scikit-learn]
- Developed collaborative filtering model using reviews (on the basis of ratings)
- Used Vector Space Modeling(Cosine Similarity) to find similar users.
- Performed matrix factorization using Low rank factorization
- Used Adam's optimization algorithm for minimizing the cost function.

Reclaiming NYC Streets [Tech stack: Pyspark, Tableau]
- Built a fare prediction model for the New York city cabs (using xgboost)
- Analyzed prospective high demand zones (using Interactive visualizations in Tableau) to recommend ways for increasing taxi ridership.
- Aggregated large-scale data (~50 M rows) from multiple sources and delivered a presentation highlighting the analysis and results.

SKILLS

- **Analytical Packages:** SQL, BigQuery ML, MS Excel, SPARK, PYSPARK, ARENA
- **Software and Programming:** Python, R, Tableau, SAS, Version Control(Git)

EXPERIENCE

AIG Analytics and Services, Bangalore, India[Tech stack: R, SQL, Tableau] — July 2014 - October 2018

Communication Model
- Identification of the most significant factors leading to customer adoption of a specified insurance product.
- "Conversion Time" windows were observed for various age groups using Tableau.
- Used ensemble of Logistic Regression and Decision tree model.

Fraud Detection Model
- Identified fraudulent transaction patterns using Random Forest and KNN classification algorithm by training on historical data.
- Used dimension reduction techniques to extract features.
- Built Logistic Regression model reducing false positives for fraudulent transactions.
- Perform data analysis and create reports by using SQL queries.

Retention Model
- Developed prediction models in R using random forest and gradient boosting to identify policies that would most likely cancel before the end of term, had 90 % True Positives.
- Performed Text mining on reviews given by customer to enable the management to know reasons for attrition.
- Analyzed customer behavioral patterns using Clustering and Decision Trees.

Validation Model
- Developed an automated validation code in SQL containing more than 150 sanity checks and was adopted by the team to be run on each new database created quarterly.

HONORS & ACTIVITIES

- *National Finalists*: Hanes Analytics Case Competition
- *Employee of the Quarter*, *consistent performer award* – EMEA and APAC, Multiple *Spot awards*
- *Volunteer*, Make a Wish Foundation
- Winner of Intramural Tennis League Singles, University of Cincinnati-2019