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

EDUCATION

University of Cincinnati, Carl H. Lindner College of Business, Cincinnati, Ohio — 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

EXPERIENCE

Circle K [Tech stack: Data Mining, PYSPARK, ML, Databricks, Time Series, Clustering]
Data Scientist — Oct 2020 - Present

- Creating metrics to measure performance of the current localized pricing model based on expected vs observed lifts
- Currently working to create a recommendation engine to specify what each store should add and drop from its shelves
- Reduced elasticity modelling timelines (six weeks to one week) through automation (using Pyspark)

AIG Analytics and Services [Tech stack: Data Mining, MSSQL, PYSPARK, ML, Random Forest, LGB, SVM, TF-IDF, Python, Snowflake, Tableau, KPI's, Stakeholder's Reporting, A/B testing]
Business Analyst (eq Data Scientist) — July 2015 - October 2018

- Responsible for owning development, involved in business analysis and technical design sessions with business and technical staff to develop requirements document and ETL design specifications
- Worked with stakeholders and other departments to convert business problems by identifying requirements and developing Machine learning models to add value and assist the business
- Provided recommendations to AIG leadership by presenting findings on key areas such as attrition, customer retention, CLV and segmentation and effectively communicate experiments, models and analytics outputs with partners

Key Projects Include:

Policy Churn Model

- Analyzing the changing attributes in a policy to figure out the features impacting the churn of a policy.
- Implemented an SVM model to predict the probability of a policy cancellation prior to the expiration date by applying TF-IDF on the various attributes of an insurance policy and analyzing the reasons for cancellation
- The final model had an accuracy, precision, recall and F-score all exceeding 95%

Propensity Model

- Identification of the potential customers having the highest probability to be converted using a gradient boosting model
- Using the same model to identify the most significant factors leading to customer adoption of a specified insurance product and performing A/B testing to prove the same
- Analyzed the future strategy based on the model insights resulting in AIG cutting down on marketing cost by 10%

Fraud Detection Model

- Identified fraudulent transaction patterns using Random Forest algorithm by training on historical claims data
- Added value to the model by reducing false positive rate, and integrating the data from the global CLUE search DB
- Helped AIG to increase the efficiency of the resource allocation in the Claims Department by 20%

Risk Analyst I — July 2014 - June 2015

- Led Datacall rollup for Specialty Line of Business (Analyzing Quarter on Quarter reports, validating treaties and other reinsurance programs and generating catastrophe reports with analysis and insights on the data)
- Developed Tableau dashboards to analyze risk portfolio using KPIs (Loss Ratio, Annual Losses, Expense Ratio, Average Policy Size) and share the same with the ERM department

PROJECTS

Optical Digit Recognition [Tech stack: Computer Vision, SVM, CNN, Python, Keras, tkinter]

- Develop a Handwritten Digits classification Model to classify a given image of a handwritten digit into one of 10 classes representing integer values from 0 to 9, inclusively
- Approach – Achieved 99.42% accuracy by performing Hyper-parameter selection and tuning for Dense Networks on CNN architecture, *and then created an API using tkinter in Python for the same*

Mercari Price Challenge [Tech stack: NLP, Python, LGBM, scikit-learn, Matplotlib]

- Develop price recommendation model for sellers of an online marketplace app to increase profitability using "best" possible selling price
- Approach – Performed data cleaning and built a gradient boosting model with hyper parameter tuning to achieve Top 100 Kaggle Public leaderboard standing with a RMSE of 0.411

Bitcoin Price Prediction System [Tech stack: Time series forecasting, R, RShiny]

- Predict future price of Bitcoin using time series and forecasting methods
- Approach – Performed EDA to get insights and differencing orders; developed a seasonal ARIMA model to forecast Bitcoin prices

SKILLS

- **Analytical Packages:** SQL, BigQuery ML, MS Excel, SPARK, PYSPARK, ARENA, Hadoop
- **Software and Programming:** Python, R, Tableau, SAS, Git, Tensorflow, Keras, GCP, Airflow, AWS, Snowflake, Azure, Looker