PIYUSH GUPTA

Minneapolis, MN 55454 • (651) 246-8778 • gupta588@umn.edu • github.com/guptapiyush340

EDUCATION

UNIVERSITY OF MINNESOTA, Carlson School of Management, Minneapolis, MN
Master of Science in Business Analytics May 2020

DR. B. R. AMBEDKAR NATIONAL INSTITUTE OF TECHNOLOGY, Jalandhar, India
Bachelor of Technology - Biotechnology May 2015

EXPERIENCE

DUNNHUMBY - TESCO, Gurgaon, India
Senior Applied Data Scientist January 2019 - May 2019
Media planning for Malaysia Suppliers

- Analyzed membership data to design 15+ media campaign for supplier in Malaysia market and measured effectiveness of targeting with A/B testing using metrics like activation and lift, leading to increased redemption rate from 2% to 3%

MU SIGMA BUSINESS SOLUTIONS, Bangalore, India
Decision Scientist October 2015 - December 2018
Benchmarking analytical platform for Bulge Bracket Investment Bank

- Identified suitable technical tools for migration by evaluating speed, scalability and accuracy of Base SAS, Microsoft R and SAS VIYA predicting loan defaulter using GBM (gradient boosting algorithms), increasing efficiency by 4 hours per model and cost savings of $2000 per year per user for Global Technology team with ~3500 analytical users

Optimization for Leading Beverage and Brewing company

- Led 12-member cross-functional team to develop machine learning based Trade Promotion optimization solution, implementing mixed integer programming and optimized investment allocation of ~$50M by maximizing sales and revenue in prototype engagement of 4 months

Predictive Analytics for American CPG manufacturer

- Designed, developed and delivered data-driven pricing simulation application based on R shiny using Monte Carlo Simulation and price elasticity model enabling strategic pricing decisions for product category with $368M net sales
- Built white box price elasticity models for 5 portfolios (80 product group) with accuracy of 85%, reducing MAPE to 15% by leveraging panel regression attributing sales to pricing, distribution and merchandising factors

Automation for Corporate Travel Management Company

- Automated 150+ reports using VBA for 10+ analyst to improve efficiency by reducing 320-person hours to 20-person

DATA SCIENCE PROJECTS

- **Data Privacy:** Developed a methodology for Fortune-10 healthcare to generate synthetic data using Generative Adversarial Network (GANs) aimed at maintaining privacy of PII data and reduced data sharing time from 12 weeks to 3 weeks
- **MinneMUDAC 2019 winner**: Predicted soybean futures contract price using LSTM & XGBoost for time series forecasting by identifying key factors like Tweets, USD index, S&P DCFI to communicate farmers to sell contracts resulting in potential savings of $7300
- **Kaggle Image classification for Dogs vs Cats**: Ranked in top 300 implementing transfer learning using CNN architecture for image classification in Keras
- **Kaggle Text Analysis for DonorsChoose.org**: Ranked in top 150 improving application screening process by predicting imbalanced class using ensemble of LightGBM and Catboost and processing textual data using TF-IDF and Word2Vec
- **Kaggle Exploratory Analysis for European soccer**: Identified relationships among winning attributes and team combination to win in Italian football league by leveraging clustering and association rules
- **Apache Airflow**: Implemented intuitive ways to schedule and monitor workflows in analytics ecosystem using apache Airflow platform which helps to track repeatable tasks and presented as part of trends in marketplace project
- **Netflix A/B Testing:** Designed a randomized-controlled experiment using thumbnails of movies to assess the impact of ethnicity on user's viewing preferences

SKILLS

- Tools: Python, R, SQL, Spark, SAS, Excel, VBA, Hadoop, Hive, Power BI, Airflow, Amazon Web Services (AWS)
- Techniques: Statistical Analysis, A/B testing, Data Visualization, Regression Analysis (Linear, Logistic, Panel, Ridge and Lasso), Decision Tree, Random Forest, XGBoost, Ensemble modelling, Naïve Bayes, KNN, Support Vector Machines, Neural Networks (RNN and CNN), Recommender Systems, Clustering, Principal Component Analysis, Optimization (Linear and Mixed Integer), Natural Language Processing (TF-IDF, Word2Vec)