Mangesh Patil

Data Science @ Pace University, NY | M.S. Graduate | mp73467n@pace.edu | +1 (516)2721857 | Portfolio | Linkedin

EDUCATION

Pace University,
MASTER'S IN DATA SCIENCE **GPA: 3.66** — **May 2024 | NYC, New York**

University of Mumbai,
BACHELOR'S IN ELECTRONICS ENGINEERING — **June 2022 | MUMBAI, INDIA**

EXPERIENCE

Maharashtra Telephone Nigam Limited | Data Analyst Intern
June 2019 – July 2020 | Thane, India

- Engineered an **RFM analysis** and **Scoring Model,** resulting in a 5% increase in targeted marketing effectiveness.
- Spearheaded Churn Prediction analysis, leading to a 3% reduction in **customer churn rate** through retention.
- Leveraged **SQL** and **Python** to **collect, clean**, and analyze large datasets, improving data-driven decision-making.
- Collaborated with the pricing team to design **personalized offers** based on RFM segmentation.

PROJECTS

Sky Data Insights: Flight Booking Analysis and Predictive Modeling | March 2024

- **Web Scraping**: Extracted and stored 1000 British Airways reviews using **BeautifulSoup**, increasing data volume for analysis.
- **Sentiment Analysis:** Cleaned data and applied **VADER**, achieving a balanced sentiment distribution **(47.6% positive, 51.9% negative**).
- Feature Engineering & Modeling: Enhanced booking dataset with new features, trained Random Forest and XGBoost models, improving **predictive accuracy to 82%.**
- Data Visualization & Reporting: Created visualizations for **sentiment** and **feature importance**, effectively communicating key insights to stakeholders.

Credit Risk Evaluation | Feb 2024

- Developed logistic regression and random forest models for credit risk evaluation, achieving **ROC-AUC** scores of 0.85 and 0.90.
- Identified key predictors of loan default—debt-to-income ratio, interest rate, and credit score—increasing prediction accuracy by 10%.
- Conducted **comparative analysis** demonstrating XGBoost's 15% improvement in precision-recall score over logistic regression for default prediction.
- Proposed refined feature engineering and parameter tuning, reducing false positives by 20% and enhancing model reliability in real-world lending scenarios.

Housing Value Navigator: Predictive Property | Dec 2023

- Python Libraries: Utilized **Pandas, NumPy, Matplotlib, and Seaborn** for streamlined data manipulation, visualization, and statistical analysis, enhancing data processing efficiency by 20% for faster insights extraction.
- Data Preprocessing: Implemented mean imputation and **IQR outlier treatment** methods, resulting in 15% improved data quality and ensuring robust data integrity for reliable analysis.
- Predictive Modeling: Developed RandomForestRegressor and TensorFlow/Keras-based deep learning models **(MLP and DNN)**, achieving an average 25% increase in predictive accuracy compared to baseline methods.
- Feature Analysis: Conducted comprehensive feature importance analysis, identifying **key predictors** of median house value and optimizing model performance with 30% reduction in complexity while maintaining or enhancing accuracy levels.

EXTRA-CURRICULAR

Cognizant's Artificial Intelligence Job Simulation on Forage| June 2023

- Performed an AI-focused job simulation for Cognizant's Data Science team, performing exploratory data analysis using **Python** and Google Colab for Gala Groceries, achieving an MAE of 0.22 (60% accuracy), and **communicated findings** through PowerPoint.

Tata Data Visualization: Business with Effective Insights Job Simulation on Forage | December 2023

- Leveraged **Tableau** to complete a simulation for Tata Consultancy Services, creating impactful **data visualizations**, formulating meeting questions for client senior leadership, and designing visuals for data analysis to empower executives in effective **decision-making**.

SKILLS

Programming Languages:
Python (Intermediate), SQL, R

Data Analysis & Visualization:
Pandas, NumPy, Matplotlib, Seaborn, Tableau, Power BI

Database Management & Querying:
MySQL, PostgreSQL, MongoDB

Statistical Analysis:
Hypothesis Testing, Regression Analysis (Linear, Logistic), Time Series Analysis, ANOVA, Bayesian Methods

Machine Learning and Deep Learning:
sci-kit-learn, TensorFlow, Keras, PyTorch, XGBoost, ARIMA, KNN, CNN, ANN, NLP

Cloud Technologies:
Microsoft Azure (AI Fundamentals, Machine Learning), Google Cloud Platform (GCP)

COURSEWORK

Data Mining | Machine Learning | Intro to Data Science | Mathematical Foundation of Analytics | Python for Data Science | Database Management and SQL | Scalable Database | Deep Learning | Algorithms | Analytical Capstone Project.

PUBLICATION

'Sales Forecasting for Telecom Vertical Using ARIMA in R' with an impact factor of 7.57/10 at IJERT Publication.

CERTIFICATION

Google Data Analytics Professional Certificate.

Microsoft Power BI Data Analytics Professional Certificate.