VIBHAV MISRA (he/him)

New York, NY 10038 | vibhavemail@gmail.com | 646-393-7014 | LinkedIn | GitHub | Portfolio

PROFILE

Data Scientist with a strong foundation in data analysis, machine learning, and artificial intelligence. Proficient in Python and adept at designing and deploying machine learning models to deliver data-driven solutions. Experienced in tackling complex challenges and collaborating with teams to implement impactful strategies across diverse industries.

EDUCATION

Pace University, Seidenberg School of Computer Science and Information Systems — **New York City, NY**
Master of Science (MS) in Data Science | **Concentration**: Data Science and Analytics — May 2026

Chandigarh University, Apex Institute of Technology — **India**
Bachelor of Engineering (BE) in Computer Science with a specialization in Artificial Intelligence and Machine Learning — June 2024

TECHNICAL SKILLS

Programming Languages: Python, SQL, R, SAS, C++, JavaScript
Machine Learning & AI: Supervised & Unsupervised Learning, Neural Networks, PCA, Clustering, KNN, LSTM, Time Series Forecasting
Data Science & Analytics Tools: Pandas, NumPy, Scikit-Learn, TensorFlow, Prophet, Keras, Jupyter Notebooks, Apache Spark, Databricks, Git/GitHub, Anaconda
Big Data & Databases: MySQL, NoSQL (MongoDB, HBase), Hadoop File System (HDFS), MapReduce
Data Visualization: Tableau, Power BI, Matplotlib, Seaborn, Excel
Additional Skills: Data Mining, Statistical Analysis (Hypothesis Testing, Regression), Data Processing (EDA, Data Cleaning)

VIRTUAL WORK EXPERIENCE

British Airways Data Science Job Simulation — **Forage - November 2024**

- Scraped and analyzed over 1,000 customer reviews using sentiment analysis techniques to uncover key insights, with 62.2% of reviews being positive and common themes centered around service, flight delays, and amenities.
- Built a predictive model using Random Forest to identify key factors influencing booking completion, achieving an accuracy of 85%.

ACADEMIC PROJECTS

Forecasting Amazon Review Ratings Time Series Analysis Using ARIMA, Prophet, and LSTM — **October 2024 – December 2024**

- Conducted time series analysis to forecast product review trends using ARIMA, Facebook Prophet, and LSTM.
- Achieved lowest RMSE of 0.55 using Prophet, outperforming ARIMA and LSTM in both accuracy and runtime.
- Engineered time-based features, handled seasonality, and visualized future sentiment trends for business use.

Electric Vehicle Adoption Analysis in Washington State — **September 2024 – November 2024**

- Analyzed EV registration trends using over 100,000 records from Washington State's open data portal.
- Identified adoption patterns by make, model year, city, and fuel type, with visualizations showing geographic and policy impact.
- Tools: Pandas, Matplotlib, Seaborn, Plotly.

Comparative Study of Decision Tree and Random Forest Models with Hyperparameter Optimization — **November 2024**

- Built and tuned classification models using GridSearchCV and RandomizedSearchCV on the Digits dataset.
- Achieved 97.5% accuracy using a tuned Random Forest, outperforming the Decision Tree model by 12%.
- Evaluated models using confusion matrices and accuracy metrics to guide model selection.

BioML — **January 2024 - April 2024**

- Developed a machine learning web application using Python and Flask to predict diseases based on user-input symptoms, achieving 90% accuracy on validation data.
- Led a team of four to build a responsive user interface using HTML, CSS, and JavaScript, integrating a predictive model utilizing Python, NumPy, and Scikit-Learn with 133 key symptom features.
- Deployed the application on GitHub, enabling real-time disease prediction.

Detecting Phishing Websites using Machine Learning — **August 2023 – November 2023**

- Built a Flask-based web application to detect phishing websites achieving an accuracy score of 85% with Gradient Boosting.
- Engineered over 30 features (e.g., URL length, domain age, HTTPS usage) to enhance model performance and designed a real-time interface for risk prediction and conducted comprehensive research on different machine learning models (SVM, Random Forest, Logistic Regression) to determine the most effective algorithm for phishing detection.

CERTIFICATIONS

- **Mathematics for Machine Learning and Data Science Specialization** (Linear Algebra for Machine Learning and Data Science, Calculus for Machine Learning and Data Science, Probability & Statistics for Machine Learning & Data Science) **Coursera – October 2023**
- **Natural Language Processing Specialization** (NLP with Classification and Vector Spaces, NLP with Probabilistic Models, NLP with Sequence Models, NLP with Attention Models) **Coursera – May 2023**
- **Applied Data Science Specialization** (Python for Data Science, AI & Development, Python Project for Data Science, Applied Data Science Capstone, Data Visualization with Python, Data Analysis with Python) **Coursera – July 2022**

ACADEMIC ACHIEVEMENTS

- Part of the pilot batch for the online BSc Degree in Programming and Data Science offered by Indian Institute of Technology Madras.
- Completed the 8 Foundational Level Courses (32 credits) and earned a Foundational Certificate from CODE, IIT Madras.
- Progressed through Diploma Level courses – Including: Machine Learning Foundations, Business Data Management, Machine Learning Techniques, Business Analytics and Tools in Data Science. Transitioned to Master's program in the USA before completing the Diploma Level.

LANGUAGES

English (Fluent), Hindi (Native)