Manisha Shah

Data Scientist

Manisa.doc@gmail.com || 360-989-0142 || Portland, OR, 97133 | LinkedIn

SKILLS

Algorithms	Machine Learning, Neural Networks, Deep Learning, NLP, Bayesian Learning, Optimization, Prediction, Pattern Identification, Data/Text Mining, Regression, Logistic Regression, Clustering, Classification, Statistical Modeling
Statistical Software	R, SAS, SPSS, MATLAB
Programming Languages	R, Python, SQL, C, C++, Java, HTML, CSS, JavaScript, Scala, NumPy, SciPy, Pandas, SciKit-learn, Matplotlib, Stats models
Data Visualization Tools	Tableau, PowerBI, Python Seaborn/Matplotlib, Excel, Power BI, and SQL Server Reporting Services.
Databases	SQL, MySQL, Hive, MongoDB, PostgreSQL
Techniques	Machine Learning, Regression, Clustering, Data Mining
Cloud Technologies	AWS, Azure
Machine Learning Experience:	Regression analysis, Naive Bayes, Decision Tree, Random Forest, LightGBM, XGBoost, Ensemble Methods, Support Vector Machine, KNN, K-Means Clustering, Natural Language Processing (NLP), Sentiment Analysis, and Neural Network., Keras, Tensor Flow

EXPERIENCE

General Electric Aviation, Portland, OR — May 2024 – Present
Data Scientist

- Developed and deployed a machine learning model to predict cost index for several Boeing and Airbus customers, reducing fuel consumption prediction error by 15%, utilizing scikit-learn, ONNX, and Python, deployed on internal FTW and ODW servers.
- Enhanced an aircraft fuel flow model by improving its sensitivity to Mach changes, resulting in a 20% accuracy increase under various flight conditions, leveraging pandas, scikit-learn, and physics-based modeling to achieve consistent performance.
- Streamlined the data processing pipeline for flight performance analysis, reducing data cleaning time by 30% through automated Python scripts and utilizing pandas and scikit-learn to efficiently handle feature engineering and preparation tasks.
- Solved a complex logic problem by calculating equivalent cost indexes for varying Mach speeds, improving operational decision-making for customers using Python and pandas, leading to better alignment with real-world flight performance.
- Created and presented a Tableau dashboard displaying real-time customer data, providing actionable insights and facilitating better decision-making for the client by integrating real-time information for more accurate operational oversight.
- Applied advanced imputation techniques, including SimpleImputer and KNNImputer, to address missing flight data, resulting in improved model robustness and accuracy, while increasing prediction reliability for operational and performance insights.
- Designed a cost index model using internal flight plan values, eliminating reliance on third-party data and enabling autonomous real-time decision-making for flight operations, utilizing Python and EMS API for seamless integration.
- Utilized tools like GridSearchCV for hyperparameter tuning, resulting in a 12% improvement in RandomForest model performance on unseen data, ensuring optimal feature selection and improving model accuracy across various flight scenarios.
- Deployed models in ONNX format, enabling integration with real-time flight systems and improving operational scalability for fuel consumption monitoring across fleets, hosted on internal FTW and ODW servers for enhanced performance.
- Conducted extensive testing and validation of the cost index model, ensuring consistent performance across multiple test scenarios and achieving a 95% success rate in validation trials, confirming model reliability for operational deployment.

Intel Corporation, Portland, OR — Jul 2023 – Sep 2023
Data Scientist

- Led an in-depth market segmentation analysis using Power BI, uncovering key trends that drove a 15% improvement in product roadmap alignment and informed strategic marketing decisions, focusing on high-growth market segments for better targeting.
- Utilized SQL for complex data extraction, transformation, and normalization, supporting CCG market models, ensuring 99% data accuracy, and improving database query efficiency by 20%, leading to more accurate insights and quicker decision-making.
- Collaborated with cross-functional teams, aligning project objectives with the Client Strategic Market Modeling team's mission to democratize data, resulting in a 30% increase in data accessibility and enhancing collaboration across different departments.
- Developed a Power BI report for the Supply Resilience team, incorporating inputs from BM, Sales, and SR teams, reducing report generation time by 25% while improving the quality of data for more informed decision-making in supply chain resilience.
- Established strong relationships with the SR core team, clearly defining roles and responsibilities, improving team cohesion by 20%, and driving the timely delivery of key project milestones to ensure the success of the initiative.
- Engaged regularly with SMEs, integrating feedback into the Rebate team BI view, which resulted in a 10% improvement in report usability and enhanced data visualization for better insights and data-driven decisions across the team.

EDUCATION

University of Texas Tech, Lubbock, Texas |Master's in data science & Business Analytics