Jay Vikas Warke

Data Scientist

+1 334-591-5807 | jay.w@ajobguide.com | Arlington, TX

SUMMARY

- Data Scientist with over 3 years of experience harnessing Python, Machine Learning, Amazon Web Services (AWS), SQL, and data management to extract valuable insights from intricate datasets.
- Skilled in Python for data manipulation, loading, and extraction, with expertise in libraries such as NumPy, Pandas, Matplotlib, Scikit-learn, Seaborn, TensorFlow, PyTorch, and NLTK.
- Proficient in developing and deploying large-scale, production-quality applications utilizing deep learning algorithms like Convolutional Neural Networks (CNNs).
- Demonstrated expertise in data visualization and statistical analysis, empowering data-driven decision-making for business intelligence initiatives through Tableau.

EDUCATION

Master of Science in Information Systems | The University of Texas at Arlington, TX (GPA:3.54)
B.Tech in Information Technology | University of Mumbai, Maharashtra, India (CGPA:7.69)

TECHNICAL SKILLS

- **Language/ IDE's:** Python, SQL, Jupyter Notebook, Google Colab
- **Machine Learning:** Linear, Logistic Regression, Decision Trees, Random Forests, Naive Bayes, SVM, A/B Testing
- **Deep Learning:** CNN, Natural Language Processing (NLP)
- **Cloud/Visualizations:** AWS, Tableau
- **Packages and Frameworks:** NumPy, Pandas, Matplotlib, Scikit-learn, Seaborn, TensorFlow, Keras, NLTK, XGBoost, PyTorch, FastAPI
- **Database:** SQL Server, MySQL, MongoDB

WORK EXPERIENCE

Allstate, TX | Data Scientist **Jan 2024 – Present**

- Developed a machine learning model to predict insurance premiums for car members, focusing on a regression problem to improve pricing accuracy.
- Utilized Python, Scikit-learn, ML algorithms (Lasso, Ridge, DNN), and regression analysis techniques for model development.
- Achieved a root mean square error (RMSE) of $28, leading to more accurate premium predictions and enhanced business profitability through better pricing strategies.
- Developed a machine learning model to detect fraudulent auto insurance claims, to reduce financial losses and streamline the claims verification process for more efficient payouts.
- Used Python (Pandas, Scikit-learn) for data preprocessing and modeling. Built a fraud detection model with Random Forest, XGBoost, and Logistic Regression, applying SMOTE for class imbalance and feature engineering (e.g., claim frequency, customer profile). Evaluated using Precision, Recall, F1 Score, and ROC-AUC.
- Identified 85% of fraudulent claims with a precision of 0.90, reducing false positives and saving 15% in fraudulent payouts, improving claims verification efficiency.
- Deployed the model using Kubernetes and Docker for scalability, integrated it with the claims system, and generated reports using SQL and Tableau for fraud pattern analysis

Trion IT Solutions, India | Data Scientist **Jan 2020 – Jul 2022**

- Conducted a detailed analysis of social media metrics and advertising campaigns to identify trends, optimize campaign performance, and support data-driven marketing decisions.
- Used Excel for social media data analysis, and Tableau for data visualization, and applied cost-benefit analysis techniques to assess campaign efficiency. Executed data cleaning and exploratory data analysis (EDA) to uncover patterns and trends.
- Improved campaign performance by 12% and enhanced data-driven decision-making processes by 25%, empowering the marketing team with actionable insights.
- Established a clustering approach to group machinery for detecting failure patterns and optimizing maintenance schedules, improving operational efficiency.
- Used Python and SQL for data extraction. Applied K-Means Clustering, DBSCAN, and PCA on sensor data (temperature, vibration) to identify maintenance patterns.
- Discovered clusters with similar failure patterns, leading to a 15% improvement in maintenance scheduling and a 10% reduction in downtime, enhancing machinery reliability.
- Deployed clustering models using Kubernetes and integrated with the maintenance system. Created Tableau dashboards to visualize clusters and key metrics for real-time insights.

CERTIFICATION

Graduate Certificate in Business Analytics