AARJAV SANGHVI
Data Scientist
312-776-6213 | aarjavsanghvi@cvmailhub.com | TX

SUMMARY

- Data Scientist with 3+ years of experience in Artificial Intelligence, Deep Learning, Machine Learning, Data Mining and Statistical Analysis.
- Expertise in building various machine learning models using algorithms such as Linear Regression, Logistic Regression, Naive Bayes, Support Vector Machines (SVM), Decision trees, KNN, K-means Clustering, and Ensemble methods (Bagging, Gradient Boosting).
- Working Knowledge of various packages in Python and R like ggplot2, pandas, NumPy, Seaborn, SciPy, Matplotlib, TensorFlow Keras and Sci-kit learn.
- Experience with high-performance computing (cluster computing on AWS with Spark/Hadoop) and building real-time analysis with Kafka and Spark Streaming.
- Proficiency in design, management and visualization of databases using MySQL, PostgreSQL and SQL Server.

TECHNICAL SKILLS

Language: Python, R, SQL, SAS
IDEs: Visual Studio Code, Anaconda
ML Algorithm: Linear Regression, Logistic Regression, Decision Trees, Classification, SVM, Random Forests, Naive Bayes, ANN, RNN, KNN, K Means, CNN
Packages: NumPy, Pandas, Matplotlib, SciPy, Scikit-learn, Seaborn, TensorFlow, Ggplot2, PyTorch, Keras, Beautiful Soup, NLTK
Cloud Technology: AWS (EC2, S3, Lambda, EMR, CloudWatch), Spark
Visualization Tools: Tableau, Power BI
Database: SQL Server, MySQL, MongoDB, PostgreSQL, SSMS
Other Tools: Git, MS Excel, Google Analytics
Operating System: Windows, Linux

EXPERIENCE

Principal Financial, IL | Data Scientist **Jan 2023 – Present**

- Implementing Classification using Supervised algorithms like Linear Regression, Logistic Regression, Decision trees, KNN, Naive Bayes
- Building comprehensive data dashboards for the effective and timely visualization of data, increasing work efficiency by 27%.
- Working with NLTK library to NLP data processing and finding the patterns.
- Collect data from end clients, performed ETL using various platforms, and defined the uniform standard format
- Improving a dashboard and story in Tableau to show the benchmarks and summary of the model's measure.
- Generating and executing a serverless architecture leveraging AWS Lambda and Amazon API Gateway, resulting in a 40% reduction in infrastructure costs and a 50% improvement in system scalability.
- Using Python (NumPy, SciPy, pandas, scikit-learn, seaborn) and Spark (PySpark, MLlib) to develop a variety of models and algorithms for analytic purposes.
- Enhance database models and querying techniques, increasing query efficiency by 22%.
- Perform univariate and multivariate analysis on the data to identify any underlying pattern in the data and associations between the variables.
- Employing Apache Spark data processing project to handle data from RDBMS and streaming sources.

XLogic Technologies, India | Data Scientist **Jan 2019 – July 2021**

- Developed various machine learning models such as Logistic regression, KNN, and Gradient Boosting with Pandas, NumPy, Seaborn, Matplotlib, and Scikit-learn in Python.
- Worked on customer segmentation using an unsupervised learning technology Performed Exploratory Data Analysis and Data Visualization using Python.
- Utilized Pandas, NumPy, SciPy, Matplotlib, Scikit-learn, and NLTK in Python for developing various machine-learning algorithms.
- Implemented statistical modeling with XGBoost machine learning software package using Python to determine the predicted probabilities of each model.
- Spearheaded initiatives to provide better and more reliable communication of data analytics to stakeholders
- Managed end-to-end complex data migration, conversion, and data modeling (using Power BI, SQL), and created visualization using Tableau to develop high-quality dashboards.
- Created and executed predictive models to improve customer acquisition, resulting in a 10% increase in new customer acquisition
- Achieved a 20% increase in overall efficiency by revamping existing queries and data models built in SQL and R.
- Participated in feature engineering such as feature intersection generating, feature normalization, and label encoding with Scikit-learn preprocessing.
- Performed preliminary data analysis using descriptive statistics and handled anomalies such as removing duplicates and imputing missing values.
- Constructed numerous data visualizations for statistical analysis and discovered meaningful data insights.

EDUCATION

Master of Science in Business Analytics **Dec 2022**
University of Illinois Chicago, IL **GPA: 4.0/4.0**

Bachelor of Engineering: Electronics and Telecommunication **Jun 2021**
University of Mumbai, India **GPA 8.23/10**