Vaishnavi V

Machine Learning Engineer

Williamsville, NY
-Email me on Indeed: http://www.indeed.com/r/Vaishnavi-V/0ece19e25ec4b150

• Result-driven professional with 6+ years of experience in Machine Learning, Data Mining with large datasets of structured and unstructured data, Data Acquision, Data Validation, Predictive Modeling and Data Visualization.
• Extensive experience in Text Analytics, developing different Statistical Machine Learning, Data Mining solutions to various business problems and generating data visualizations using R, Python, and Tableau.
• Rich exposure in NLP with Apache, Hadoop and Python
• Hands on SparkMlib utilities such as Including Classification, Regression, Clustering, Collaborative Filtering, Dimensionality Reduction.
• Proficient in Statistical Modelling and Machine Learning techniques (Liner, Logistics, Decision Trees, Random Forest, SVM, K - Nearest Neighbours, Bayesian, XG Boost) in Forecasting/Predictive Analytics, Segmentation Methodologies, Regression based Models, Hypothesis Testing, Factor Analysis/ PCA, Ensembles.
• Skills in implementing LDA, Naïve Bayes, Neural Networks, Principal Component Analysis, and Recommender Systems.
• Expertise in transforming business requirements into analytical models, designing algorithms, building models, developing data mining and reporting solutions that scales across massive volume of structured and unstructured data.
• Adept in statistical programming languages like R and Python including Big data technologies like Hadoop, Hive.
• In-depth knowledge in Software Development Life Cycle (SDLC) including Requirement Analysis, Design specification, and Testing as per cycle in both waterfall and agile methodologies.
• Skilled in using dplyr and pandas in R and Python for performing exploratory data analysis.
• Experience in designing stunning visualizations using Tableau software, and publishing and presenting dashboards, Storyline on web and desktop platforms.
• Experience in designing and developing the Tableau and updating the existing desktop, developing ad-hoc reports, scheduling the processes and administering and tableau activities using tableau.
• Prediction - prediction of numerical value using Regression or CART
• Experience with Data Analytics, Data Reporting, Ad-hoc Reporting, Graphs, Scales, Pivot Tables, and OLAP reporting.
• Highly skilled in using visualization tools like Tableau and ggplot2 for creating dashboards.
• Worked and extracted data from various database sources such as SQL Server, DB2, S3 Bucket, and JIRA (internal issue trackers for the project development).
• Well experienced in Normalization & De-Normalization techniques for optimum performance in relational and dimensional database environments.

Authorized to work in the US for any employer

Work Experience

Machine Learning Engineer

Customers Bank - Phoenixville, PA
January 2021 to Present

Responsibilities:
• Evaluated research needs and used Time series for efficient statistical programming and Analysis.
• Developed Generalized Linear Models, Stepwise Regression, Classification, and Regression Trees.
• Used RNNs for Time Series Forecasting. Developed LSTM and GRU forecasting models.
• Designed and implemented Tableau visualization solutions.
• Collaborate with research coordinators and statisticians to develop a statistical analysis plan.
• Analyze data using statistical software to generate descriptive statistics and to develop statistical models and graphics.
• Review statistical output for consistency and quality.
• Communicate statistical results to the manager.
• Defined data rules for technical requirements of customer accounts and built KPIs for performance metrics.
• Implemented Content based Filtering in recommender system to customize the offers to the debit and credit card users.

Environment: Python 3.5.2, Tableau 9.03, R, GIT, LINUX, , Machine Learning, Recommender System, timetk, keras, Tensorflow.

Machine Learning Engineer

FOURKITES - Chicago, IL
August 2017 to August 2020

Responsibilities:

• A highly immersive Data Science program involving Data Manipulations & Visualization, Web scraping, Machine Learning, Python programming, SQL, GIT, Linux commands, No SQL, Mongo DB, Hadoop.
• Designed and developed Tableau Reports, Documents, Dashboards for specific requirements and timelines.
• Storage and data analysis tools setup in Amazon Web Services cloud computing infrastructure.
• Used Pandas, numPy, seaborn, scipy, matplotlib, scikit-learn, NLTK in python for developing various Machine Learning algorithms.
• Prepared Dashboards using Calculations, parameters, in Tableau.
• Participated in all phases of data mining, data collection, data cleaning, developing models, validations, visualizations, and performed gap analysis.
• Data Manipulation and aggregations from different sources.
• Programmed a utility in python that used multiple packages (scipy, numpy, pandas).
• Implemented classification using supervised algorithms like Logistic Regression, Decision Trees, KNN, Naïve Bayes.
• Good knowledge of Hadoop architecture and various components such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node, Secondary Name Node, and Map Reduce concepts.
• Updated Python scripts to match training data with database stored in AWS cloud search, so that we would be able to assign each document a response label for further classification.
• Data Transformation from various resources, data organization, feature extraction from raw and stored.

• Validated the machine Learning classifiers using ROC curves and Lift Charts.
Highlights:
• Designed Dynamic Asset Assignment System to match trucks to loads dynamically by inferring from the location trajectory of trucks using Siamese Networks which improved tracking consistency and supply chain visibility.
• Forecasted future truck locations in case of inconsistent location updates, based on lane behaviors using Generative LSTM models and enhanced the customer experience

Environment: Tableau 9.03, GIT, LINUX, Python 3.5.2, Machine Learning, MLLib, Regression, Logistic Regression, Hadoop NoSQL, random forest, HDFS, SVM, Map Reduce.

Data Analyst

ATHENA HEALTHCARE - Watertown, MA
December 2013 to July 2017

Responsibilities:

• Performed Data Cleaning, feature scaling, feature engineering using pandas and numPy packages in Python.
• Performed data cleaning and imputation of missing data.
• Used clustering technique K-Means to identify outliers and to classify unlabeled data.
• Principal Component Analysis is used in feature engineering to analyze high dimensionality data.
• Exploring DAG's and their dependencies using Air Flow Pipelines for job automation.
• Ensured that the model has low False Positive Rate and Text classification for unstructured and semi-structured data.
• Overfitting was addressed by implementing the regularization methods like L1 and L2.
• Tracking jobs until certain criteria is met using Air Flow Technology.
• Created Data Quality scripts using SQL to validate successful data load and quality of the data.
• Designed and built a text classification application using different text classification models.
• Created customized business reports and shared insights to the management.
• Learned new tools and skillsets as per the need.
• Preparing associated documentation for specifications, requirements, and testing.
• Communicated the results with operations team for taking best decisions.
• Ensure the code is deployed properly in the production environment.
• Optimize the data pipeline for CICD (Continuous Integration and Continuous Deployment).
• Experienced in Agile Methodologies and SCRUM process.
• Used JIRA for defect tracking and project management
Highlights:
• Designed AI Chatbot to converse with the patients and other care providers to gather basic information while the system was busy finding the specialists to answer the queries.
• Designed auto Debugging Chatbot to make troubleshooting easy for the operations team to fix the customer complaints and to make the developers life easy to find the root cause for the issue.
Environment: Python 2.x, GIT, LINUX, Python 3.5.2, Machine Learning, Regression, Logistic Regression, Hadoop 2.3, MySQL, random forest, NLTK, SVM, KMeans, Keras, TensorFlow, Jenkins.

Education

Master of Science in Engineering Sciences
University at Buffalo - SUNY - Buffalo, NY
August 2020 to Present

Bachelor of Engineering in Electronics and Communication Engineering
JNTU University - Ananthapur, AP, IN
August 2009 to June 2013

Skills

- BI Tools: Tableau, Tableau server, Tableau Reader, Amazon Redshift.
- Reporting Tools: MS Office, Tableau.
- Version Control Tools: GitHub, Perforce.
- Regression Polynomial Regression
- Random Forest
- Logistic Regression
- Decision Trees
- Classification
- Clustering
- Association
- Simple/Multiple linear
- Kernel SVM
- K Nearest Neighbors (KNN).
- Hadoop
- Hive
- HDFS
- Map Reduce
- Kafka.
- ggplot2
- caret
- dplyr
- gmodels
- RCurl
- twitter
- NLP
- Reshape2
- pandas
- numPy
- seaborn

- sciPy
- matplotlib scikit-learn
- Theano
- TensorFlow
- sqlalchemy.
- SQL
- Python
- R
- MATLAB. PERL
- SCALA
- JAVA
- HTML.
- SQL
- Mongo DB
- Elastic Search
- HDFS
- MySQL
- Cassandra
- Teradata.
- Windows
- Linux
- Unix
- Macintosh HD.
- Spark