**Thribhuvan Sai Teja** +1-571-250-0445

**Data Scientist** Vivek@tektroop.com

**SUMMARY:**

* Data Scientist / Machine Learning Engineer with 8 years of experience in handling Structured and Unstructured data, writing advanced SQL queries, Da ta Wrangling, Data Visualization, Data Acquisition, Predictive modeling, Probabilistic Graphical Models, Inferential statistics, Data Validation.
* Experience in building robust Machine Learning, Deep Learning and Natural Language Processing Models.
* Expertise in Statistical analysis, Text mining, Supervised learning, Unsupervised Learning, and Reinforcement learning.
* Proven experience working on Supervised, Unsupervised techniques such as Regression, Classification, Clustering, Machine Learning (ML), Deep Learning (DL).
* Experience in writing Big data queries using Apache Spark with Python and Scala (Spark RDD and Data frames) to pull data from different sources, perform Big data analytics and build Machine Learning models using SparkMLlib.
* Experience with Python Libraries Pandas, NumPy, Seaborn, Matplotlib, NLTK, Spacy, Scikit - learn, Keras and TensorFlow in developing end to end Analytics and ML, DNN models.
* Experience in plotting visuals, building dashboards and storytelling using Tableau, AWS Quick Sight, Matplotlib, Seaborn, Plotly and Power BI.
* Strong mathematical background in Linear algebra, Probability, Statistics, Differentiation and Integration.
* Strong skills in statistical methodologies such as A/B test, experiment design, hypothesis test, ANOVA, and Time Series Analysis.
* Hands-on experience on Azure Cloud and ML API process via Flask.
* Having deep understanding of state-of-the-art machine learning and deep learning algorithms, techniques and best practices.
* Proficient in Machine Learning techniques (Decision Trees, Linear/Logistic Regressors, Random Forest, SVM, Bayesian, XG Boost, K-Nearest Neighbors) and Statistical Modeling in Forecasting/ Predictive Analytics, Segmentation methodologies, Regression based models, Hypothesis testing, Factor analysis/ PCA, Ensembles.
* Strong experience in Text Mining of cleaning and manipulating text and finding sentiment analysis from text mining data.
* Hands on experience with Big Data tools like Hive, Pig, Sqoop, Apache Flume and Kafka.
* Experience in building models with TensorFlow and top-level frameworks such as Keras, Theano and Pytorch.
* Good knowledge of Hadoop architecture and its components like HDFS, MapReduce, Job Tracker, Task Tracker, Name Node and Data Node.
* Experience in designing Star schema, Snow flake schema for Data Warehouse, ODS architecture.
* Skilled in Advanced Regression Modeling, Correlation, Multivariate Analysis, Model Building, Business Intelligence tools and application of Statistical Concepts.
* Excellent understanding of Software Development Life Cycle (SDLC), Agile, Scrum and waterfall.
* Experience working in AWS environment using S3, Athena, Lambda, AWS Sage maker, AWS Lex, AWS Aurora, Quick Sight, Cloud formation, Cloud Watch, IAM, Glacier, EC2, EMR, Recognition and API Gateway.
* Good industry knowledge, analytical & problem-solving skills and ability to work well with in a team as well as an individual.

**TECHNICAL SKILLS:**

**Programming Languages / Software libraries:**Python, R, Java, Scala, SQL, TensorFlow

**Supervised and Unsupervised:**XGBoost, Light GBM, Artificial Neural Networks, Auto encoders, Convolutional Neural Networks, Recurrent Neural Networks, LSTM, Bi-directional LSTM, ANN, CNN, Multi-Layer perceptron's, Linear Regression, Polynomial Regression, Logistic Regression, SVM, Random Forests, Decision Trees, K-NN, Naive Bayes, K-Means, Hierarchical clustering, Association Rule Learning, Reinforcement Learning, Self-organizing maps

**Dimensionality Reduction Techniques**Principle Component Analysis (PCA), Latent Dirichlet Allocation (LDA), Kernel PCA.

**Model Evaluation / Engineering:**Cross Validation Technique, Activation Functions, Grid Search, Bayesian Optimization and Regularization (Lasso and Ridge Regression), Feature Selection methods, Feature Scaling.

**Natural Language Processing (NLP):**Text Analytics, Text processing (Tokenization, Lemmatization), Text Classification, Text clustering, Name Entity Recognition (NER), Word Embedding and Word2Vec, POS Tagging, Speech Analytics, Sentimental Analysis.

**Python Programming Skills:**Keras, Pandas, Numpy, scikit-learn, NLTK, SpaCy, SciPy, PySpark, Plotly, Cufflinks, Seaborn, Theano, matplotlib, Django, Flask, GloVe, Pytorch, Beautiful Soap (bs4), Web Scraping

**R Programming Skills:**R Shiny, MICE, rpart, CARET, random Forest, Data Preprocessing, Web Scraping, Data Extraction, Dplyr, GGplot2, Statistical Analysis, Predictive Analysis, GGplotly, rvest, Data visualization

**Data Visualization:**AWS Quick sight, Tableau, MS Power BI, Seaborn, Qlik View matplotlib, Plotly, cufflinks, ggplot2, RShiny.

**Big Data:**Hadoop, Hive, Mongo DB, Apache Spark, Scala, Pig, Sqoop

**Database Servers:**MySQL, Microsoft SQL server, SQLite, Red Shift, PostgreSQL, Mongo DB, Teradata

**Amazon Web Services:**EC2, Lambda, Sage Maker, EMR, S3, Quick Sight, API Gateway, Athena, Lex, Recognition, CI/CD, Code Commit, DynamoDB, transcribe, Cloud Formation, Cloud Watch, Glacier, IAM

**Development Environments/ Cloud:**AWS, IBM Cloud, Azure

**PROFESSIONAL EXPERIENCE:**

**T Mobile, Seattle WA(Remote) Dec 2021 - Present**

**Data Scientist**

**Responsibilities:**

* Prepare and Analyze data, identify the patterns on dataset by applying models and statistical analysis. Collaborating with Senior Data Scientists for understanding data.
* Experience on Migrating SQL database to Azure data Lake, Azure data lake Analytics, Azure SQL Database, Data Bricks and Azure SQL Data warehouse and controlling and granting database access and Migrating On premise databases to Azure Data lake store using Azure Data factory.
* Develop GUI using JQuery Json and Java script.
* Perform data preparation, data manipulation, normalization, and predictive modelling. Improve accuracy by evaluating model in Python and R.
* Perform Sensitive Analysis and Topic Classification on open-ended surveys by text mining to help with business improvements.
* Perform customer segmentation based on machine learning and statistical modelling effort including building predictive models and generate data products to support customer segmentation.
* Used Python and R for programming for improvement of model. Upgrade the entire models for improvement of the product.
* Under supervision of Sr. Data Scientist performed Data Transformation method for Re scaling and Normalizing Variables
* Developed Spark applications using Pyspark and Spark-SQL for data extraction, transformation and aggregation from multiple file formats for analyzing & transforming the data to uncover insights into the customer usage patterns.
* Performed K-means clustering, Multivariate analysis and Support Vector Machines.
* Worked on Natural Language Processing with NLTK module for application development for automated customer response.
* Utilized machine learning algorithms such as linear regression, multivariate regression, Naive Bayes, Random Forests, K-means, & KNN for data analysis.
* Advanced Text analytics using Deep learning techniques such as Convolution neural networks to determine the sentiment of texts.
* Supported client by developing Machine Learning Algorithms on Big Data using PySpark to analyze transaction fraud, Cluster Analysis etc.
* Used Spark for test data analytics using MLLib and Analyzed the performance to identify bottlenecks.
* Converted the unstructured data into structured data using Apache Avro.
* Designed predictive models using the machine learning platform -H2O, Flow UI.
* Used the Agile Scrum methodology to build the different phases of Software development life cycle.
* Built multi-layers Neural Networks to implement Deep Learning by using Tensor flow and Keras.
* Worked with Machine learning algorithms like Regressions (linear, logistic), SVMs and Decision trees.
* Used Text Mining and NLP techniques find the sentiment about the organization.
* Developed MapReduce/Spark modules for machine learning & predictive analytics in Hadoop on AWS.
* Utilized various new supervised and unsupervised machine learning algorithms/software to perform NLP tasks and compare performances.
* Design and develop analytics, machine learning models, and visualizations that drive performance and provide insights, from prototyping to production deployment and product recommendation and allocation planning.
* Utilized Spark MLlib, SparkSQL, Python, R, Databricks a broad variety of machine learning methods including classifications, regressions, dimensionality reduction, etc.
* Worked with sales and Marketing team for Partner and collaborate with a cross-functional team to frame and answer important data questions prototyping and experimentation ML algorithms and integrating into production system for different business needs.
* Worked on Multiple datasets containing two billion values which are structured and unstructured data about web applications usage and online customer surveys.
* Performed Data cleaning process applied Backward - Forward filling methods on dataset for handling missing values.
* Segmented the customers based on demographics using K-means Clustering.
* Presented Dashboards to Higher Management for more Insights using Tableau.
* Used classification techniques including Logistic Regression, Decision Tree to predict the likelihood of each user to churn.
* Performed Boosting method on predicted model for the improve efficiency of the model.
* Collaborating with the project managers and business owners to understand their organizational processes and help design the necessary reports

**Technology Stack:**Python, Spark (PySpark, SparkSQL), R/R Studio, MS SQL Server, Tableau, AWS, MS Excel, Microsoft Outlook

**BCBS, Jacksonville FL Aug 2018- Dec 2021**

**Data Scientist/Machine Learning Engineer**

**Responsibilities:**

* Participated in all phases of data mining, data collection, data cleaning, developing models, validation, and visualization to deliver data science solutions.
* Built a model to predict direct mail marketing campaign for loan pre-approvals to optimize the number of letters being sent.
* Created robust Java Script GUI applications using Object-Oriented programming.
* Worked on fraud detection analysis on payments transactions using the history of transactions with supervised learning methods.
* Collected data using database tool to retrieve the data required for building models such as MS SQL Server.
* Experience working with Azure Databricks, Azure SQL database, Azure SQL Datawarehouse
* Used statistical methods and data mining methods in Python to explore the fast growth opportunities of the clients.
* Clustering customers by RFM analysis using Python, selected customers with high value and improved their retention rate by sending ads and coupons.
* Worked Closely with different Product owners to get more business insights.
* Expert knowledge on Trend Lines, Statistics, and Log Axes.
* Worked on Amazon Web Services cloud virtual machine to do machine learning on big data.
* Implemented a Python-based distributed random forest via PySpark.
* Used Pandas, Numpy, Seaborn, Matplotlib, Scikit-learn in Python for developing various machine learning models and utilized algorithms such as Linear regression, Logistic regression, Gradient Boosting, SVM and KNN
* Cleaned data by analyzing and eliminating duplicate and inaccurate data using R.
* Experience in retrieving unstructured data from different sites such as in html, xml format.
* Performing Exploratory Data Analysis on the data provided by the Client.
* Worked with Pandas libraries in R for storing the retrieved data in a file.
* Conducting Root Cause Analysis independently/collaboratively in resolving multiple occurred issues.
* Worked with Data frames and other data interfaces in R for retrieving and storing the data.
* Responsible in making sure that the data is accurate with no outliers.
* Applied various machine learning algorithms K-Means in R with the required packages installed.
* Applied K-Means algorithm in determining the position of an Agent based on the data collected.
* Applied Regression in identifying the probability of the Agent's location regarding the insurance policies sold.
* Used advanced Microsoft Excel functions such as Pivot tables and VLOOKUP in order to analyze the data and prepare programs.
* Performed various statistical tests for clear understanding to the client.
* Actively involved in Analysis, Development and Unit testing of the data and delivery assurance of the user story in Agile Environment.
* Worked with Quality Control Teams to develop Test Plan and Test Cases.
* Super Intended usage of open source tools - R Studio(R) and for statistical analysis and building the machine learning models.
* Worked with different data science teams and provided respective data as required on an ad-hoc request basis
* Assisted both application engineering and data scientist teams in mutual Agreements/provisions of data.
* Used cross-validation to test the models with different batches of data to optimize the models and prevent overfitting.
* Used PCA and other feature engineering techniques for high dimensional datasets while maintaining the variance of most important features.
* Created Transformation Pipelines for preprocessing large amount of data with methods such as imputing, scaling, selecting and etc.
* Performed Ensemble methods to increase the accuracy of the training model with different Bagging and Boosting methods.

**Technology Stack:**PySpark, Python 3.x (Numpy, Pandas, Scikit-learn, Matplotlib), Jupyter, MS SQL Server, AWS, SparkSQL, AWS, Tableau 9.x, Jupyter Notebooks

**Caterpillar Inc. - Chicago, IL Nov 2015 – Aug 2017**

**Sr. Data Analyst/Data Scientist**

**Responsibilities**

* Worked closely with data scientists to assist on feature engineering, model training frameworks, and model deployments implementing documentation discipline.
* Involved with Data Analysis primarily Identifying Data Sets, Source Data, Source Meta Data, Data Definitions and Data Formats
* Performed Decision Tree Analysis and Random forests for strategic planning and forecasting and manipulating and cleaning data using dplyr and tidyr packages in Python.
* Involved in data analysis and creating data mapping documents to capture source to target transformation rules.
* Creating or modifying the T-SQL queries as per the business requirements and worked on creating role playing dimensions, fact-less Fact, snowflake and star schemas.
* Wrote, executed, performance tuned SQL Queries for Data Analysis & Profiling and wrote complex SQL queries using joins, sub queries and correlated sub queries.
* Involved in development and implementation of SSIS, SSRS and SSAS application solutions for various business units across the organization.
* Developed mappings to load Fact and Dimension tables, SCD Type 1 and SCD Type 2 dimensions and Incremental loading and unit tested the mappings.
* Wrote test cases, developed Test scripts using SQL and PL/SQL for UAT.
* Worked with the ETL team to document the Transformation Rules for Data Migration from OLTP to Warehouse Environment for reporting purposes.
* Performed data testing, tested ETL mappings (Transformation logic), tested stored procedures, and tested the XML messages.
* Created Use cases, activity report, logical components to extract business process flows and workflows involved in the project using Rational Rose, UML and Microsoft Visio.
* Wrote complex SQL queries for validating the data against different kinds of reports generated by Business Objects XIR2.
* Developed Data Mapping, Transformation and Cleansing rules for the Master Data Management Architecture involved OLTP, ODS and OLAP.
* Extensively used SQL, T-SQL and PL/SQL to write stored procedures, functions, packages and triggers.
* Analyzed of data report were prepared weekly, biweekly, monthly using MS Excel, SQL & Unix.
* Applied various machine learning algorithms and statistical modeling like decision tree, logistic regression, Gradient Boosting Machine to build predictive model using scikit-learn package in Python.

**Environment:** Python, T-SQL, SSIS, SSRS, SQL, PL/SQL, OLTP, Oracle, MS Access, MS Excel, XML, Microsoft Visio, UML, OLAP, Unix