Veda Narayana  
**Data Scientist/Analytics /AI-ML Engineer**   
Email: Vedanarayanareddychintalapudi@gmail.com  
Contact: 5715313053

* 10+ years of experience inData Science and Analytics includingData collection, Data Cleaning, Data Manipulation, Data Mining, Data Validation, Data Visualization, Analytics modeling and Designing algorithms
* Extensively worked on **Python**and **R**with various packages, **MATLAB** with various toolboxes
* Solid ability to write and optimize diverse **SQL** queries, and working knowledge of RDBMS and NoSQL DBMS
* Encountered a wide range of data format including Structured data (Relational, Non-Relational), Semi-structured data, Unstructured data
* Strong Knowledge in Bid Data environment and using tools like **Hive**and cloud-based data warehouse such as **Teradata, Snowflake, Amazon Redshift, Amazon S3, Google Bigquery, Azure synapse Analytics,**
* Responsible for implementing Map Reduce programs into Spark transformations using Spark and Scala.
* Fluency in Python with working knowledge of ML & Statistical libraries (e.g. Scikit-learn, Pandas).
* ITindustry. Experience includes Network security using ArcSight, Wireshark, Sidewinder, SourceFire and BlueCoat. I possess excellent problem solving, analyze, interpersonal, and follow-through skills. Also, I am detail oriented, multi-tasking with strong organizational abilities and able to work with all level of staff and clients.
* GoodExperience with a focus onBig data, Deep Learning, Machine Learning, Image processing or Al.
* Experience in processing real-timedata and building ML pipelines end to end.
* Resolve all failures related to the PKI Certificate Authority infrastructure
* Implemented Spring boot microservices to process the messages into the Kafka cluster setup.
* Developed Spark Applications by using Scala, Java and Implemented Apache Spark data processing project to handle data from various RDBMS and Streaming sources.
* Proficient in using different AI/ML platformslike**Anaconda (JupyterNotenook, Spyder), Google Colab,TensorFlow, Alteryx, Dataiku, RapidMiner, KNIME, Google Vertex AI &AutoML**for data pre-processing and model prediction
* Very good hands - on in Spark Core, Spark SQL, Spark Streaming and Spark machine learning using Scala and Python programming languages.
* Operate, maintain, manage, and upgrade the entire agency PKI/PIV infrastructure.
* Analyze Packet captures using wireshark and Netscout.
* Supporting users of Final Cut Pro edit systems and users of the NBA Media Archive
* Experienced in**MLOps**Concepts (Deploying, Monitoring, Maintenance)using **AmazonSagemaker, Databricks, Google Cloud AI platform**
* Getting real time data using Kafka and processing using Spark and Scala.
* Developed Spark Programs using Scala and Java APl's and performed transformations and actions on DD's.
* Assisting with NBA Media Archive account management/setup for staff; troubleshooting all issues related to the DMM work flows.
* Good experience in extracting and analyzing the very large volume of data covering a wide range of information from a user profile to transaction history using machine learning tools.
* Closely worked with Kafka Admin team to set up Kafka cluster setup on the QA and Production environments.
* Directly support the Confidential implementation of DoD PKI with operational focus on the implementation, management and sustainment of the Marine Enterprise PKI. Work directly on the continued implementation of the DoD PKI within the Confidential, both NIPR and SIR, to include enterprise certificate validation infrastructure, directory services and support to deployed forces. Responsible for the management of USMC PKI Infrastructure as it relates to use of DoD PKI and CAC and
* SIPR Token.
* Extensively used Python requests urllib and urllib2 module to call web services.
* Used **Git 2.X, SVN**for version control and **Apache Maven/Ant**for build
* Passionate about gleaning insightful information from massive data assets and developing a culture of sound, data-driven decision making
* Developed scripts in Python and Excel VA to automate the data analysis, generating statistics isolating trends in memory failures.
* Implemented Bagging and Boosting to enhance the model performance
* Highly skilled in statistical methodologies includesExperimental design and hypothesis test such as ANOVA test, T test, F test, Chi-square test, A/B testto prove the business assumption through data-driven analysis using **Python** and **Excel**
* Familiarity with machine learning algorithm such asLinear Regression, Logistic Regression, Decision Tree, Random Forest, Naive Bayes, KNN, SVM, Neural Networks, CNN, RNN, K-means and Hierarchical clustering
* Hands on experience in building sophisticated models to address the business problems with machine learning models, Segmentation, Time series and other advanced analytical/statistical methodologies in **Retail, Health care, Telecom, Insurance** and**financial**(**Bank, Credit card, Loan)**services
* Performed data analysis and applied ML models in Various field includes**medical diagnosis**, **product recommendation**, **fraud detection**, **Speech recognition**, **Image recognition** and **Stock market trading**
* Experienced in ML model testing, Model deployment, model monitoring and model maintenance
* Excellent communication skills and understanding of System Development Life Cycle (SDLC) with Agile,Scrum and waterfall methodology
* Successfully working in fast-paced multitasking environment both independently and in collaborative team, a self-motivated enthusiastic learner

**Technical Skills:**

|  |  |
| --- | --- |
| Languages | Python, R, SQL |
| Packages | **Python** - Pandas, NumPy, TensorFlow, Seaborn, NLP(NLTK), SciPy, Matplotlib, Scikit-learn, Beautiful Soup, Scrapy, Rpy2, Keras  **R** - ggplot2, caret, rvest, dplyr, gmodels,NLP (tm), Reshape2, rjson |
| Cloud Platform (Datawarehouse) | Snowflake, Teradata, AWS (Redshift, S3), GCP- Google Bigquery, Microsoft Azure (Azure synapse Analytics), IBM DB2 |
| Databases | MySQL, SQL Server, Oracle, NoSQL-MongoDB |
| Data Analytical/Statistical/ Build models Platform | MS Excel, RapidMiner, KNIME, Dataiku, Alteryx, Google Cloud Vertex AI &AutoML, Amazon Sagemaker |
| BI Tools (Reporting, Analyzing and Visualization) | Tableau, Power BI, QlikView, MS Office (Word/Excel/Power Point/Visio) |
| ML Ops (Deploying, Monitoring, Maintenance) | Amazon Sagemaker, Databricks, Google Cloud AI platform |
| Data Modelling Tools | MS Visio |
| Data Processing Framework | Hadoop (HDFS), Spark |
| Version Control Tools | SVN, GitHub |
| Project Execution | Rational Unified Process (RUP), Rapid Application Development (RAD) |
| Operating System | Windows, Linux, Unix |

**Macy’s, New York Jan 23 to present.**

**Title: Data Scientist/AI-ML Engineer**

**Responsibilities:**

* Developed Python modules for day-to-day business activities, incorporating machine learning and predictive analytics.
* Used multiplemachine learning algorithms, including random forest and boosted tree, SVM, SGD, neural network, and deep learning using TensorFlow.
* Utilized the Gamma-Gamma model as an extension to the Pareto model, with a specific focus on purchase count, lifetime, and monetary value.
* Employed forecasting models like ARIMA and Holt-Winter for various tasks assigned by higher management.
* Converted data from PDF to XML using python script in two ways i.e. from raw xmI to processed ×mi and from processed xml too.CSV files.
* Responsible for providing operational and technical support for the Media Operations Group.
* Analyze network traffic PCAP and Tcpdump with wireshark.
* Had knowledge on Kibana and Elastic search to identify the Kafka message failure scenarios.
* Worked on loading CSV/TXT/AVRO/PARQUET files using Scala/Java language in Spark Framework and process the data by creating Spark Data frame and RDD and save the file in parquet format in HDFS to load into fact table using ORC Reader.
* Utilized different behavior tracking algorithms to predict customer bank loan default probabilities.
* Modify PKI Group Policy Object settings and update scripts for helpdesk personnel when necessary to remediate certificate-related issues.
* An excellent understanding of both traditional statistical modeling and Machine Learning techniques and algorithms like Regression, clustering, ensembling (random forest, gradient boosting), deep learning (neural networks), etc.
* Develop ETL Process usingSPARK, SCALA, HIVE and HBASE.
* Loaded the data into Spark DD and do in memory data Computation to generate the Output response.
* Used python Element Tree(ET) to parse through the XML which is derived from PDF files.
* Build digital workflows and ensure that all content processed by the 24/7 Media Operations Group is prepped, QC'ed, and fulfilled on time and in the proper specification for international distribution.
* Narrow down anomaly traffic with wireshark for hostile string or Domain.
* Used ScalaTest for writing test cases and coordinated with QA team on end to end testing.
* Expertise python scripting with focus on DevOps tools, CI/CD and AWS Cloud Architecture.
* Implemented a range of machine learning algorithms to analyze customer satisfaction reports and devised various customer retention strategies.
* Implemented to reprocess the failure messages in Kafka using offset id.
* Developed and implemented Customer Churn models to assess and gain insights into customer churn, enabling effective retention strategies.
* Developed highly scalable classifiers and tools by leveraging machine learning, Apache spark & deep learning
* Participate in the implementation and management of all SAN and VM Infrastructure associated with USMC PKI
* Good knowledge in setting up batch intervals, split intervals and window intervals in Spark Streaming.
* Utilized Market Mix Modeling to strategize advertisement investments and optimize ROI on advertisements.
* Worked under the direction of (SO to develop an effective solution to a predictive analytics problem, testing a number of potential machine learning algorithms of apachespark.
* Implemented several research proof-of-concept models for real-time fraud detection in credit card and online banking transactions.
* Developing software in Python, C# and web development frameworks such as Flask
* Utilized Credit Analysis and Risk modeling algorithms to implement customer acquisition strategies in real-time business scenarios.
* Assisted in building data modeling and ETL pipelines such as extracting, Transforming and storing in GCP- Google BigQuery and Amazon Redshiftdata warehouse using Alteryx
* Designed, built, and deployed machine learning models using AWS SageMaker in a production-ready hosted environment.
* Implemented Customer Survival Analysis models to predict when customers are likely to return for service.
* Created a daily support knowledge base to streamline troubleshooting mechanisms.
* Studied and implemented fraud detection models to monitor unconventional customer purchases and provide timely alerts.
* Built artificial neural networks and worked on Natural Language Processing (NLP), Speech recognition, and computer vision projects.
* Worked in GCP- Google Bigqueryand Snowflake data warehouse for data storage and management
* Designed neural networks using TensorFlow-Keras API, applying rectifier and sigmoid functions, and compiled entire Artificial Neural Network (ANN) into a classifier to predict customer churn rate.
* Trained the model through 100 epochs, dividing them into 10 batches, using Stochastic Gradient Descent, achieving an accuracy of 83.6%.
* Experienced in developing REST APIs and deploying ML models using CI/CD tools like Jenkins.
* Worked on creating Docker containers to efficiently manage the application lifecycle.

**Environment:** Machine learning, Python, Jupiter, Pytorch, PySpark, Flask web framework, AWSsage maker, MS Azure, Cassandra, SAS, Apache Spark, HDFS, Hive, Pig, Linux, Anaconda Python, MySQL, Eclipse, SQL connector, Spark ML, Jenkins, Dockers, Kubernetes, Umap, Numpy, A/B Testing, R-Shiny, inference testing.

**Pfizer, New York May 21 to Dec 22**

**Data Scientist / Analytics**

**Responsibilities:**

* Involved in complete Software Development Life Cycle (SDLC) process by analyzing business requirements and understanding the functional workflow of information from source systems to destination systems
* Assisted in building data modeling and **ETL** pipelines such as extracting, Transforming and storing in **GCP- Google BigQuery** and **Amazon Redshift**data warehouse using **Alteryx**
* Performed customed**data mapping**, **integration** and transformation process using **Apache Spark**
* Analyze PKI and PIV-I requirements, policies, and procedures to help define viable courses of action to integrate PIV-I smart
* cards into the Confidential enterprise.
* Monitor all work order tracking tools to ensure on - time delivery using the Media Broker and Web Makelist, OTTO, and DMD success notifications, among others. If content is at risk of not being delivered resolve the issue or escalate appropriately
* Trace frames or packets to analyze network-related issues with protocol analyzers such as Wireshark, NI observer, Network General Sniffer or Cisco NAM.
* Used Spring Kafka API calls to process the messages smoothly on Kafka Cluster setup.
* Used Scala collection framework to store and process the complex consumer information.
* Imported real time weblogs using Kafka as a messaging system and ingested the data to Spark Streaming.
* Used the XML tags and attributes to isolate headings,side-headings,and subheadings to each row in CSV file.
* Analyzed and Prepared data, identify the patterns on dataset by applying historical models
* Mentored highly immersive Data Analytics program involving Data fetching, Data manipulation and preparation, data profiling, data mining,Visualization and Machine Learning models using **SQL** and **Python**libraries in **Anaconda**(Jupyter notebook) platforms
* Developed test scripts in Python that are used to test embedded systems via their debug UART port.
* Assist with the coordination all PKI and PIV-I requirements across multiple internal and external stakeholders.
* Provide 24/7 support for the Media Operations Center - including serving as a point of escalation for content or workflow concerns and troubleshooting issues related to Avid files, Ingest, Tape Deliverables, Encoding systems and Edit Rooms
* Good experience in extracting and analyzingthe very large volume of data covering a wide range of information from a user profile to transaction history using machine learning tools.
* Implemented Kafka producer and consumer applications on Kafka cluster setup with help of Zookeeper.
* Converted mahout's machine learning algorithms to RDD based ApacheSparkMLLib to improve performance.
* Used Scala functional programming concepts to develop business logic.
* Experienced in using Spark Core for joining the data do deliver the reports and for detecting the fraudulent activities.
* Provided training for Application of various statistical distribution and machine learning algorithms like**Logical Regression**, **Decision Trees**, **Random Forest**, **neural networks**, **SVM**, **clustering, K-means**to identify Volume using scikit-learn package
* Worked and Trained people in using **Alteryx**, **RapidMiner** for data preparation, exploration and analysis, visualization and predict modeling
* Mentored juniors for developing and deploying machine learning models using **Google cloud Vertex AI** and **Google AutoML**
* Used various algorithms of PySparkMLAPI.
* Use Wireshark on a consistent basis to resolve connectivity problems between hosts
* Very good hands-on in Spark Core, Spark SQL, Spark Streaming and Spark machine learning using Scala and Python programming languages.
* Have knowledge on partition of Kafka messages and setting up the replication factors in Kafka Cluster.
* Experienced with Spark Context, Spark-SQL, Data Frame, Datasets, Spark YARN.
* Developed Views and Templates with Python and to create a user-friendly website interface Django's view controller and template.
* Used **Python** and **R** for developing various machine learning algorithms to predict the data, and Improved efficiency and accuracy by evaluating modelto forecast the data for better results
* Documented project reporting requirements and translate them into functional specifications
* Predicted user preference based on segmentation using General Additive Models, combined with feature clustering, to understand non-linear patterns between user segmentation and related monthly platform usage features (time series forecasting)
* Performed Boosting method on predicted model for the improve efficiency of the model
* Developed highly scalable classifiers and tools by leveraging machine learning, Apache spark & deeplearning.
* Implemented Key Performance Indicator (KPI) Objects, Actions, Hierarchies and Attribute Relationships to measure the performance of an organization, department and project
* Worked on Agile Methodologies (JIRA) and presentedvarious analytical Dashboards to Higher Management for more Insights using **Power BI** and **Tableau**
* Knowledge of model maintenance techniques, including updating and retraining models as new data becomes available, or as the data distribution changes
* Collaborated with operation team and developer in MLOps environment, that is deployment, monitoring, and maintenance using Google Could AIor Databricks

**Environment:**(Descriptive Analysis, Predictive & Prescriptive Analysis, Exploratory Analysis)

**DXC Technology Co - Conway, AR. Mar 19 to Apr 21**

**Role: Data Scientist**

**Responsibilities:**

* Developed monthly, quarterly, and annual reports based on claims, exceptions, renewal, and other attributes in Web Intelligence with aggregate awareness and prompt reports to enhance 10% report performance.
* Utilize Wireshark, FireEye Web MPS and Nitro SE|M tools to investigate, respond to, and generate reports for GNOSC and CyberComm security incident tickets
* Used Unit Test library for testing python and other programs.
* Performed Statistical Modelling with ML to bring Insights in Data under the guidance of the Principal Data Scientist.
* Worked with the procurement domain and built various dashboards for different modules that showed yearly, quarterly, and monthly operations of the cockpit. Added scripts for bookmark, filter, page navigation, OpenDoc, hyper linking, reduced dashboard loading time 20%.
* Performed data wrangling to clean, transform and reshape the data utilizing panda’s library.
* Analyzed data using SQL, R, Java, Scala, Python, Apache Spark and presented analytical reports to management and technical teams.
* Worked under the direction of CSO to develop an effective solution to a predictive analytics problem, testing a number of
* potential machine learning algorithms of apache spark.
* Worked with different datasets which includes both structured and unstructured data and
* Participated in all phases of Data mining, Data cleaning, Data collection, variable selection, feature engineering, developing models, Validation and Visualization.
* Developed and implemented predictive models of user behavior data on websites, URL categorical, social network analysis, social mining and search content based on large-scale MachineLearning.
* Worked on different Machine Learning models like Logistic Regression, Multi-layer perceptionclassifier and K-means clustering.
* Used R and Python for Exploratory Data Analysis to compare and identify the effectiveness of the data.
* Implemented applications with Scala along with Akka and Play framework.
* Created clusters to classify control and test groups.
* Analyzed and calculated the life cost of everyone in a welfare system using 20 years of historical data.
* Expert in implementing advanced procedures like text analytics and processing using the in-memory computing capabilities like Apache Spark written in Scala.
* Used Python, R, SQL to create statistical algorithms involving Multivariate Regression, Linear Regression, Logistic Regression, PCA, Random Forest models, Decision trees, SVM for estimating and identifying the risks of welfare dependency.
* Extensively used Pandas, NumPy, Seaborn, Matplotlib, Scikit-learn, SciPy and NLTK in R for developing various machine learning algorithms.

**Environment:** SQL/Server, Oracle, MS-Office, Teradata, Informatica, ER Studio, XML, Business

Objects, HDFS, Teradata, JSON, HADOOP (HDFS), MapReduce, PIG, Spark, R Studio, MAHOUT, JAVA, HIVE, AWS.

**BlueCross BlueShield (BCBS) - Chicago, IL.**

**Data Scientist Nov 17 to Feb 19**

**Responsibilities:**

* Involved in creating technical design documents out of requirement documents.
* Used multiple machine learning algorithms, including random forest and boosted tree, SVM, SGD, neural network, and deep learning using TensorFlow.
* Identify the key variables, perform the required data cleaning, data manipulations, and data preparation.
* Understanding the data and preparing the necessary inputs to the data.
* Documents collection, preprocessing, Tokenization.
* Developed and implemented predictive models of user behavior data on websites, URL categorical, social network analysis, social mining and search content based on large-scale Machine Learning,
* Removing Stop words, stemming words, indexing. Detailed effort spending classification against support type, rework type, line of business, team, location, activity type, etc.
* Applied confusion matrix to find the accuracy threshold value.
* Train and evaluate Models using SVM, DECISION TREES, and NAÏVE BAYES.

**Environment:** R, Python, NLP (Text Analytics)

**SiriusXM- Lawrenceville, NJ. Jan 16 to Oct 17**

**Role: Data Analyst.**

**Responsibilities:**

* Proficient in understanding and analyzing data of the financial domain, Meetings with stakeholders to understand and set business problems.
* Extracted data from various sources which include flat files, and databases such as Oracle using SQL.
* Analyzed the data for data gaps and performed Exploratory data analysis on the data by performing univariate and bivariate analysis on the data with the help of pandas
* Analyzed the data for identifying the gaps in the data, analyzing the various patterns of financial transactions of the data.
* Using clustering algorithms such as K means classifying the transactions and analyzing the clusters and creating the profiles for clustering to classify the clusters by labeling the clusters.
* Identified anomalous/fraudulent transactions using the isolation forest algorithm to mark the transaction for further investigation.
* Based on the type of transaction, transaction amount, reason, type of account, frequency of transactions, location, time, age, balance, rolling balance, etc., to check for fraudulent transactions and created a dashboard using PowerBI.
* For version control git was used.
* Deployment of analysis results and created data pipelines to classify future transactions by using AWS EC2 environment exposing flask, Docker.

**Environment:** AWS (EC2, ECS), K Means, Flask, git, PowerBI, isolation forest, Oracle SQL, EDA, Univariate and Bivariate Analysis), Python, Pandas, NumPy, scikit-learn.

**Qualcomm, Hyderabad, India Jul 13 to May 15**

**Role: Sr. Data Analyst.**

**Responsibilities:**

* Designed and developed a call quality control system using Python and machine learning techniques such as Random Forest regressor using Scikit-learn.
* Utilized the Random Forest Regressor algorithm to create a highly accurate model for predicting call quality metrics.
* Conducted feature selection using GridSearchCV to identify and select the most important features for predicting call quality.
* Conducted data cleaning and preprocessing using Pandas and NumPy to ensure the data were suitable for analysis.
* Used standardization techniques to normalize the data and ensure consistency across all call quality metrics.
* Developed and utilized Scikit-Learn modeling pipelines to automate the entire modeling process and increase efficiency.
* Utilized the SciPy library for statistical analysis and hypothesis testing to identify potential correlations and causations between call quality metrics.
* Conducted regular testing and maintenance to ensure the call quality control system was always running optimally.
* Worked with a large dataset of call records to analyze and predict call quality metrics for both voice and data calls.
* Used data visualization tools by using matplotlib to present insights and trends in the call quality data to stakeholders.
* Conducted data exploration to identify potential outliers and anomalies in the call quality data.
* Created documentation and training materials to enable stakeholders to understand and use the call quality control system effectively.
* Developed and implemented automated alerts to quickly identify and resolve any issues with call quality.

**Environment:** Python, Scikit-learn, pandas, matplotlib, SciPy, NumPy, GridSearchCV, Random Forest.