Ram Reddy – Data Scientist / ML Engineer

**BOSTON, MA** | +1 559-493-0177 | [vramireddy45@gmail.com](mailto:vramireddy45@gmail.com)

[linkedin.com/in/venkata-rami-reddy-kallu-030a1b77](https://www.linkedin.com/in/venkata-rami-reddy-kallu-030a1b77)

Professional Summary:

* Expert in bridging the gap between tech and business working with multiple clients to turn requirements without an obvious blueprint into scalable modeling solutions using techniques that cut across entire technical stack.
* Demonstrated track record of hands-on coding experience in Python, R, and SQL.
* Ability to analyze raw data, drawing conclusions and developing recommendations.
* Work experience in analytics, working with data to convert large volumes of structured and unstructured data into actionable insights and business values.
* Proficient in Machine Learning, Data/Text Mining, Statistical Analysis & Predictive Modeling.
* Efficient in data acquisition, storage, analysis, integration, predictive modeling, logistic regression, decision ​trees, data mining methods, forecasting, factor analysis, cluster analysis, ANOVA and other advanced statistical​ techniques​.
* Experience in statistical data analysis like Chi-square, T-test; Dimensionality reduction methods like PCA, LDA and feature selection methods.
* Strong experience and knowledge in Data Visualization with Matplotlib and Seaborn: Line and scatter plots, Bar​ ​Charts, Histograms, Pie chart, Box plots, subplots etc.
* Experience in designing, developing, scheduling reports/dashboards.
* Experience in data mining & loading and analyzing unstructured data -XML, JSON, flat file formats.
* Experienced in solving Classification and Regression problems using Machine Learning algorithms such as Linear Regression, Logistic Regression, Decision Trees, Random Forests, Support Vector Machines, Naïve Bayes etc., and optimizing the performance using Ensemble Learning techniques like Bagging, Boosting and Slacking.
* Proficient in evaluating and analyzing the performance of model with Cross validation, ROC-AUC, Confusion Matrix, Precision, Recall, Log-Loss, MSPE, MSAE, R Square, F1-Score and creating Performance reports using tables and visualizations.
* Highly motivated team, effective communication, analytical and presentations player with excellent interpersonal skills.

Technical Skills & Capabilities:

|  |  |
| --- | --- |
| **Programming Languages** | Python (NumPy, SciPy, Pandas, Scikit- Learn, Matplotlib,  Statsmodels, Seaborn, Graph Lab, NLTK, bokeh, SpaCy, Dash,Plotly,Clips, Jupyter​ ​Notebook) |
| **Databases** | MySQL, PostgreSQL, MongoDB |
| **Big Data Technologies** | Hadoop (Map Reduce), Spark (SparkSQL, SparkML), Hive,Pig,HBase,HiveQL |
| **Cloud Technologies** | AWS (Sage Maker-Studio, Feature Store, S3, EC2), Cloudera |
| **Operating Systems** | Unix, Windows |
| **Classification** | Naive Bayes, SVM |
| **Regression** | Linear Regression, Logistic Regression, Penalized Linear​ ​regression |
| **Ensemble** | Random Forest, Boosting |
| **Instance-based** | k-Nearest Neighbor (kNN) |
| **Decision Tree Learning** | Classification and Regression Tree (CART), Gradient Boosting Machines (GBM) |
| **Clustering** | K-Means, Hierarchical, DBScan |
| **Recommendation Engines** | Content-Based filtering, Collaborative filtering |
| **Text Analytics** | Stemming, NLTK, Pandas, TFIDF Vectorizer, Word Cloud, Topic Modelling |

# Experience:

**State Street Corporation, Boston, MA May 2017 to Present**

ML Engineer/ MLOps Engineer

* Performing Statistical Modelling, Data Extraction, Data Screening & Cleaning, Data Exploration, Feature Selection, Feature Engineering and Data Visualization, Linear and Logistic Modelling, Dimensionality Reduction and implementing Machine Learning algorithms on large scale to deliver resourceful insights and inferences targeted towards boosting revenues and enriching customer experiences.
* Worked with text feature engineering techniques n-grams, TF-IDF, word2vec etc. Applied various Classification models such as Naïve Bayes, Logistic Regression, Random Forests, Support Vector Classifiers, Stochastic Gradient Descent, RNN, LSTM from scikit-learn and Keras library.
* Performed Confusion Matrix and Classification report to evaluate accuracy and performance of different models used. Evaluated the model’s performance using various metrics like Precision, Recall, F1-Score, AUC-ROC, Cross Validation to test the models with different batches of data to optimize models.
* Performed Feature Engineering such as Missing Value Imputation, Normalization and Scaling, Outliers Detection and Treatment, One-Hot-Encoding, Splitting Features and used Label Encoder to convert categorical variables to numerical values using python scikit-learn library.
* Analyzing data using a variety of advanced Analytics and Machine Learning techniques to build an initial proof of concept supporting the hypothesis.
* Building quality prototypes and implementing the solutions in a scalable framework using Cloudera Spark.
* Developing high-performance algorithms for precision targeting, testing, and implementing these algorithms in scalable, production-ready code.
* Applying traditional as well as modern statistical techniques, including Regression, Support Vector Machines, Regularization, Boosting, Random Forests and other Ensemble Methods.
* Developed Deep Neural Network models for classification using Tensor Flow.
* Train, Test, Evaluate and Deploy analytical models.

**Environment:** ​Mac OS, Windows, ​​Python 3.x, AWS Sage Maker pipelines, TFX, Cloudera Data Science workbench, Agile.

Data Scientist

* Worked closely with internal stakeholders such as business teams, product managers to meet the company's reporting and analytical needs.
* Performed Data cleansing/transformation, feature engineering, apply statistical analysis, data visualization, data mining.
* Established scalable, efficient, automated processes for model (predictive/machine learning) development, model validation, model implementation and data analysis.
* Predicted and monitored fund volume and alerted the business when there is an unusual activity.
* Handled importing data from various data sources, performed transformations using Hive, Spark and loaded data into HDFS.
* Draw inferences and conclusions, and create visualizations of processed data, identify trends, anomalies, and present the results to the business stakeholders.
* Worked with very large number of data sets (both structured and unstructured) and designed, developed and implemented R&D and pre-production prototype solutions.
* Leveraged advanced Machine Learning techniques to discover and solve new business problems.
* Identified process improvements that significantly reduce workloads or improve quality.
* Develop predictive analytics for funds processed through various process steps enabling business to understand their performance issues.
* Collaborated with data engineering teams to productionize the analytical models.
* Provided ongoing tracking and monitoring of performance of decision systems and statistical models.
* Developed real time predictive models using machine learning techniques like Gradient descent, XG-Boost, etc. for predicting volume of funds and tickers processed by business during the critical processing time window of 4 to 7 pm.
* Implemented clustering algorithms to group the clients with similar behavior.
* Used the Spark Python API (PySpark) to expose the spark programming model developed in python so that the code be run on the BDE
* Ensure there is proper sync between the model and the UI so that visualizations can be shown
* Configure spark streaming to the data pulled through the source database so that model can be invoked, and desired output/results can be generated.
* Configure hive tables and store the model output into hive tables so that the same can be accessed by the UI service layer.

**Environment:** ​Mac OS, Windows, ​​Python (and related libraries),​ ​Linux​, ​SQL​, ​Hadoop, Spark (PySpark), Hive, Cloudera Data Science workbench, Agile

# Education and Certifications

## DeepLearning.ai - Coursera

[Machine Learning Engineering for Production (MLOps) Specialization.](https://www.coursera.org/account/accomplishments/specialization/VFQWXTVBQT58) **[Apr 2022 – Oct 2022](https://www.coursera.org/account/accomplishments/specialization/VFQWXTVBQT58)**

[Practical Data Science on the AWS Cloud Specialization](https://www.coursera.org/account/accomplishments/specialization/4CCLA7XL7WMK) **[Jun 2022 – Oct 2022](https://www.coursera.org/account/accomplishments/specialization/4CCLA7XL7WMK)**

**San Francisco Bay University,** *Fremont, CA* **Aug 2015 – Aug 2016**

Master of Science in Computer Science

## Institute For Technology and Management, Mumbai, *India.* Jul 2013 – Jul 2015

Post Graduate Diploma in Supply Chain and Operations Management

## Narasaraopet Engineering College, AP, *India.* Jun 2007 – Jul 2011

Bachelor of Technology in Computer Science and Engineering