**.V.Dora Reddy**

Mobile: +91 – 8341226699

Email: vvdorareddy6666@gmail.com

**Professional Summary:**

* Overall 6.8 years of rich experience in ETL Testing, Bigdata and Machine Learning.
* **Executing Data-driven Solutions** to increase efficiency, accuracy, and utility of internal data processing. Experienced at creating data **Regression models**, using **predictive data modeling, and analyzing data mining algorithms to deliver insights and implement action-oriented solutions to complex business problems**.
* Over 2+ years of experience in **Machine Learning** (*Predictive Analytics, Regression techniques,* ***Linear*** *Regression,* ***Logistic*** *Regression, Decision Tree,* ***SVM****,* ***Naive Bayes****, KNN, K-Means,* ***Ensemble*** *Models, Random Forest, Dimensionality Reduction Algorithms* ***PCA****,* ***Recommendations****, Gradient* ***Boosting*** *algorithms (GBM &* ***XGBoost****)),*

**Deep Learning** (Artificial Neural Networks, Convolutional Neural Networks, Recurrent Neural Networks and LSTM).

* Performed Data Analysis, **Statistical Model Building**, Predictive Analysis, **Parameter Tuning, Visualization in Python**
* Rich Knowledge in **Statistics and Probability**
* Creation of automation scripts for repetitive tasks to save the development hours and improve the standards of the coding and testing activities.
* Over 2+years of experience in **Big Data Technologies**-**Hadoop**, **HIVE**, **PIG**, **SQOOP**, **HBASE**, **KAFKA**, **OOZIE**, **SPARK**.
* Extensive experience of **SQL** plus, Toad for designing SQL statements.
* Worked on importing data from Oracle, MySql, Teradata into HDFS and Hive using Sqoop, TPT.

**Professional Experience:**

# Currently working as Machine learning Engineer at Root Infosoft Technology Private Limited, Hyderabad from Feb’12 to till date

**Technical Summary:**

**Machine Learning: Supervised Learning** & **Unsupervised Learning:**

Linear regression, Logistic regression, Decision trees, Random Forest, Boosting Algorithms (Gradient Boosting Machine & XG Boosting) Naive Bayes, Support Vector Machines, Clustering (K-means), Principal component analysis (PCA), K-nearest neighbors (KNN),

Natural Language Processing (NLP), **Deep learning** --- ANN,CNN,RNN .

**Hadoop/Big Data/NoSQL:**  HDFS, MapReduce, Hive, Pig, Sqoop, HBase, Spark Sql, Kafka , Spark core, Spark ML.

**Programing Languages** : C, C++, Core Java, **Python** & **R**, Scala, Spark .

**Database** : Oracle, MySQL, HDFS, HBASE.

**Version Control Tools** : GitHub

**Reporting Tools** : Tableau

**Academic Qualification:**

* **B.Tech** (Civil Engineering) with 81% from V.R.Siddhartha Engineering College in 2008
* **Intermediate** with 96.2% from Githanjali College in 2003
* **10thClass(CBSE)** with 86% from Triplaar School Of Learning in 2001

**Project #1**

**Project :  DAVRS**

**Client** : Mtouch Labs pvt ltd, Hyderabad.

**Technologies** **involved**: Machine Learning, Python(sklearn), Flask, Zappa, Deep Learning.

**Database**: Aws

**Repository**: Github

**Duration:** Jun’17- till Date

Dubshoot Audio Video recommendation systems (DAVRS) is a two-model based system. One model which uses the numerical ratings of similar items (audio and video files) between the active user and other users of the Dubshoot app to assess the similarity between users’ profiles to predict recommendations of unseen items to both active and other users.

Other model which uses the most popular items (audio and video files) Dubbed by the users and recommend N- top or popular items to users.

**Responsibilities:**

* Develop and build data models for use in applications and process improvements
* Building machine learning frameworks to drive personalisation and recommendations
* Leverage off-the-shelf solutions to develop customized data visualizations
* Develop expertise in concepts, methods and supporting technology relevant to analytics
* Research & develop Machine Learning models for Recommendation problems, in the areas of Audio, Application & Data
* Provide SW specifications, production quality code and engage with algorithm proliferation activities
* Research and develop deep learning algorithms like CNN, RNN, ANN, LSTM

**Project #2**

**Project:** **Management Awareness System**

**Client:** City of Detroit, MI, USA

**Duration:** Sep’16 – Feb ’18 (in MACHINE LEARNING Project)

**Duration:** Apr ’12 – Mar ’14 (in PL/SQL Project)

**Description:** The Management Awareness System (MAS) is the process of systematic analysis of a Detroit Police Department member’s performance patterns to alert supervisors of the possible need for intervention or special recognition. This System is a thorough, reliable and extensible platform for the acquisition, analysis, reporting and monitoring of early warning behavioral characteristics of members of Detroit Police Department.

**Responsibilities:**

* Experience in Analysis, Design, Maintenance, Support and Coding using Oracle, PL/SQL and worked on various phases of projects including development, enhancement, maintenance and support on Client Server with hands-on experience in tools like TOAD and SQL Developer.
* Well conversant in translating business requirements into the creation of database objects: tables, indexes, views, constraints, packages, stored procedures, functions etc.
* Lead analytic and technical innovation through framework and model development
* Develop algorithms and build data models for use in applications and process improvements
* Building machine learning frameworks to drive personalisation and recommendations
* Leverage off-the-shelf solutions to develop customized data visualizations
* Develop expertise in concepts, methods and supporting technology relevant to analytics

**Project #3**

**Project** **: Migration &**

**Client : Nissan, CA**

**Duration** **:** Jun’14 to Nov’16

**Environment** **:** HADOOP, HDFS, SQOOP, HIVE, SPARK, SCALA, KAFKA, Ubuntu

**Description:**

The project involved performing Data Ingestion from various DataBases Into

HDFS and also use of HIVE QUERY LANGUAGE (HQL) to create appropriate Hive tables,

loading the structured data resulted from MapReduce jobs into the tables and writing

hive queries to further analyze the logs to identify issues and future Analytics purposes.

**Roles & Responsibilities:**

* Importing data from RDBMS to HDFS and HIVE using SQOOP.
* Collecting and aggregating large amounts of log data using Apache Flume or Kafka and staging data in HDFS for further analysis.
* Creating many UDF’s with complex types as input and outputs for analysis and aggregations.
* Experienced in working with SPARK eco system using SCALA and HIVE Queries on different data formats like Text file and parquet.
* Experienced in working with Partitions, Bucketing concepts in Hive and designed both Managed and External tables in Hive to optimize performance.
* Code review as per the customer coding standards.

**(V.V. Dora Reddy)**