Vijayakumar Subramani

Email: vijaysubramani.inbox1@gmail.com

My Works: <https://github.com/sravanvvijay> Mob: 7338800153

**Career Objective**

Over 11+ years of total experience in development and support applications and have 3 years’ experience in Machine learning and Data Science. I wish to seek challenging assignments and responsibilities, with an opportunity for growth and career advancement as successful achievements through a continuous learning process and keep myself dynamic, visionary and competitive with the changing scenario of the world.

**Professional**

 Hands-on experience in Supervised Machine Learning algorithms like Linear Regression, Logistic Regression, Naïve Bayes, KNearest Neighbor, Decision Tress, Random Forest and Ensemble Models

 Hands-on experience in Unsupervised Machine Learning algorithms like Kmeans, Hierarchical Clustering, Dimension Reduction Techniques

 Hands-on Experience in R, Python, ANSI SQL, Big Data technologies in Windows and UNIX environment.

 Good Hands on in Python, Pandas/Numpy, Scikit Learn and NLTK Libraries

**Technical Skills**

 **Software Products:** R, Python, Hadoop **-** Hive, Spark, HDFS, Apache Kafka, Sqoop and Cloudera

 **Databases:** Oracle, DB2, and MySQL

 **Operating System:** Windows, UNIX

 **Applications:** MS Office

 **Cloud Technologies:** AWS and GCP

**Certifications**.

 Full Stack Data Science from Jigsaw Academy

 AWS Certified Solution Architect-Associate

 CCA175- Cloudera Hadoop and SPARK developer

**Projects Profile**

**Project Title: Telecom Churn Prediction**

**Project details & Responsibilities:**

 Predicted Customer Attrition in Telecom sector using classification techniques. Identified Target Segment for Proactive Customer Retention Campaigns and Revenue Maximization.

 Created Data Quality Report

 Variable Profiling for Continuous and Categorical Variables

 Data Preparation: Binning, Missing Value Imputation, outlier detection, Derived Variables and Dummy Variable creation

Model Building: used stepwise regression model using R Programming Language

 Applied different classification model algorithms

 Final Model selection

 Created customer segments for high, medium and low profile customers

**Project Title: Predict a Click**

**Problem Statement:** One of the famous hotel accommodation-booking agency. The objective of the project is to anticipate how likely users are to click on a specific hotel and how often a hotel will be clicked based on certain characteristics of the hotel

**Project details & Responsibilities:**

 Predicted user interest using regression techniques, which determines what hotels and advertisers we present to the user and at what position they will be shown

 Created Data Quality Report

 Data Preprocessing

 EDA & Data Preparation:

Decile Binning

Missing Value Imputation

Outlier detection

Derived Variables

Dummy Variable creation

 Model Building: used all regression techniques and Hyper Tuning Parameters

Liner Regression (Used Shrinkage Methods – Lasso and Ridge)

Polynomial Regression

Random Forest

 Final Model selection

**Project Title: Provider Fraud Detection**

**Problem Statement:** Predicts fraud in the medical insurance industry using anomaly analysis and geo-demographic metrics. Insurance providers will enable Pulse insights gain credibility in the industry, combat the increasing costs of healthcare, and costly impact of fraud.

**Project details & Responsibilities:**

 Used Google Cloud Big query API to load the data into GCP

 Created Data Quality Report

 Data Exploration, Cleaning and Preprocessing

 Data Preparation

 built a data model to build the dataset from Multiple CMS data source

 Feature Engineering

 Model Building: used classification Techniques

Logistics Regression

Random Forest

 Optimization Purpose: used boosting Model

**Project Title: New York Cab Fare Prediction**

**Problem Statement:** Predicting the fare amount for a taxi ride in New York City given the pickup and drop off locations

**Project details & Responsibilities:**

 Created VM in GCP using system configuration 24 GB RAM, 3 Core CPU as the dataset has 50 million observations

 Used Kaggle API to load the data into GCP VM

 Hypothesis Generation – Analyzed factor that generally affect the cab fare

 Data Sampling

 Data Exploration, Cleaning and Preprocessing

 Data Preparation

 Feature Engineering

 Model Building: used Regression Techniques

Liner Regression (Used Shrinkage Methods – Lasso and Ridge)

Polynomial Regression

Random Forest

**Project Title: Credit Risk Modeling**

**Problem Statement:** Predicting the credit risk by using the techniques expected loss, probability of default, loss given default and exposure at default

**Project details & Responsibilities:**

 Hypothesis Generation

 General Preprocessing

 PD Model: Data Preparation

 PD Model Estimation

 PD Model Validation

 LGD and EAD Models

This is to acknowledge that the above provided information is true and correct to the best of my knowledge. Any misrepresentation of facts or data will be entirely liable upon me.

**Place:** Chennai

**Date:** 28th May, 2020 Vijayakumar Subramani