**Venkatesh Kalla**

**305-521-9404**

venkate7sh@gmail.com

<https://www.linkedin.com/in/venkateshkalla/>

**Summary:** Data scientist with 5.5 years of experience in utilizing machine learning, data processing and data analysis skills to solve challenging analytical problems and aid in business decision making.

**Skills & Technologies**

|  |  |
| --- | --- |
| Certifications: | Data Analyst Nano Degree from Udacity, Neural Networks and Deep Learning from deeplearning.ai, Java SE6 Programmer, AWS Developer Associate and Solution Architect |
| Programming: | Python, R, SAS, SQL, Java, XML, Shell, Spark, Scala, Pig, Hive, Map Reduce |
| Software: | Anaconda 3.5, Tableau, AWS, SAS Enterprise Miner, R Studio, Postman API |
| Competencies: | Machine Learning, Business Intelligence, Business Analysis, Data Mining, Data Engineering, Statistical Analysis, Time Series Analysis, Forecasting, Anova, A/B Testing, SDLC, Waterfall and Agile |

**Education**

|  |  |
| --- | --- |
| **Fairleigh Dickinson University, Teaneck, NJ, USA** | August 2014 – December 2015 |
| Masters in Computer Science |  |
| **Jawaharlal Nehru Technological University, Kakinada, India** | August 2008 – July 2012 |
| Bachelors in Electronics and Communication Engineering |  |

**Work Experience**

***Barclays, Data Scientist, Wilmington, DE*  May 2016 – Present**

**Responsibilities:**

* Designing and developing various machine learning algorithms for POC (Proof of Concept) and fraud data pooling
* Working on POC’s with Apache Spark APIs over Cloudera Hadoop YARN, used Scala for data processing and performed data analytics on Hive DB
* Customizing core libraries to be used by all the team members in the Docker AWS-based analytics environment with other standardized python and R dependencies
* Implementing python-based machine learning models as RESTful web services and consuming them using REST APIs
* Used random forest algorithm to build a fraud detection model that improved the detection rate of customer fraud by 10%
* Used natural language processing toolkit and VADER package for performing sentiment analysis on online customer complaints and identify negative sentiment scores that improved model performance by 15% than traditional model

***Sprint, Data Scientist, Overland Park, KS* June 2015 – April 2016**

**Responsibilities:**

* Built a churn prediction model using logistic regression that predicted subscription cancellations and aided business in designing personalized campaigns for the users that reduced churn rate approximately by 4%
* Developed various machine learning models for business in the areas of campaign management that demonstrated performances of 80.6% on par with state-of-the-art models used in industry
* Used random forest algorithm to help analyze various acquisition campaigns and improve their effectiveness approximately by 80%
* Used Spark Streaming APIs to build a common learner data model for near real-time data processing that collects data from Kafka and persists it into Cassandra

***Syntel, Data Analyst, Chennai, India* June 2012 – August 2014**

**Responsibilities:**

* Built a multi-class classification algorithm that accurately classified credit card holders and improved bank’s efficiency by reducing default rate and offering new product choices to customers
* Performed credit scoring analysis to predict whether or not to extend credit limits for a new or an existing applicant and if it results in a profit or loss
* Developed various data visualizations using Tableau to analyze data trends over time
* Used logistic regression, clustering and multivariate modeling to provide valuable analytical insights
* Used Kolmogorov-Smirnov test (K-S test or KS test) to measure the quality of the models, performed feature visualization and used k-fold cross validation to avoid model over fitting

**Course Work**

* Improved policy renewal rate by an average of 23% by building a churn prediction model using logistic regression in SAS Enterprise Miner
* Built a cost optimized model to predict Expedia bookings with an accuracy of 0.84 and an average loss of $0.27/$1 using ensemble methods
* Analyzed the purchasing behavior of customers using a customized analytics model in base SAS to predict who preferred Barnes & Noble over Amazon
* Developed an ER and a Dimensional model for a website through reverse engineering by identifying the key entities of the business process using SQL server
* Learnt to solve many data mining challenges through Kaggle projects: <https://github.com/utdkaggleprojects/>
* Built various visualizations on Tableau Public: https://public.tableau.com/profile/venkateshkalla#!/