**Vivekkumar Singh**

14305 Wedgewood Ct Tampa, FL- 33613 |Phone: 8136062975 | [viveks2692@gmail.com](mailto:viveks2692@gmail.com) | [Linkedin](https://www.linkedin.com/in/vivekkumarusf2020/) | [Github](https://github.com/Vivekkumar260892): Vivekkumar260892

**SUMMARY**

Over 4 years of comprehensive work experience in Data Science, Marketing Analytics and Business Intelligence in banking, retail and supply chain domains. Ability to solve complex business problems using ETL, Data Warehousing, Machine Learning and Exploratory Data Analysis.

**EDUCATION**

* MS in Business Analytics and Information Systems, University of South Florida ***(Jan 2019-May2020)***
* BE in Computer Engineering, University of Mumbai ***(Jul 2011- Jul 2015)***
* Certifications**:** [Google Digital Unlocked](https://drive.google.com/open?id=1hk-yppx7N8wtPaYNUsswoy-JoMxn-ktL), [AWS certified Cloud Practitioner](https://www.youracclaim.com/earner/earned/badge/916e6481-cbff-4165-982c-d460209cd2e1), [Tableau Data Scientist](https://www.youracclaim.com/badges/8e401080-061f-4a25-a041-c27f5f333b1f)

**WORK EXPERIENCE**

***Assistant Decision Sciences at University of South Florida: (Jun 2019- May 2020)***

* Performed Data **Cleaning**, Data **Manipulation** and Data **Integration** to maintain Data Quality for creating various assignments.
* Created **Tableau** dashboards to explain variation in success **Metrics** and **Time Series Analysis** toundergraduate students.
* Automated assignment grading process using **Excel VBA (Macros)** and **MySQL** maintaining accuracy and saving **~ 75%** of time.

***Analyst at Cartesian Consulting: (Mar 2018- Dec 2018*)**

* Predicted sales by **time series forecasting** in **Python** using **neural networks, ARIMAX** and **Prophet** for inventory management by eliminating understocking and reducing overstocking by **56%**.
* Applied **K-means clustering** in **Python** for **segmentation** of customers comparing it with **RFM** (Recency, Frequency and Monetary Value) analysis for improved **CRM** campaign targeting.
* Gathered requirements and developed pipelines for **ETL** (Extract, Transform, Load) using **SQL** ,**Python** and **Airflow**.

**•** Optimized customer channel preference model without any business drop, **reducing** the campaign **deployment** **costs** by **30%.**

* Generated visualizations using **Tableau** tointerpret data, analyze marketing **metrics** and **supply** **chain** analysis.
* Created **KPIs** to build customer one view. Performed **hypothesis** **tests** and **statistical** **analysis** between variables.
* Developed **dimensional** **data** **models** and **data** **warehouse** adhering to **integrity** and **normalization** rules to support **CRM** data mart. Wrote **complex** **SQL** queries (multiple table joins, CASE WHEN, CTE’s, subqueries) to extract valuable insights.

***Associate Analyst at Hansa Cequity:* (*Mar 2016- Sep 2017*)**

* Identified probable customer churn using **Predictive Models** in **Python** like **Logistic** **Regression**, **Decision** **Trees**, **Random Forest** and achieved a true positive rate (recall) of **84**% for target customer retention and acquisition marketing campaigns.
* Validated **ROC curves**, **confusion** **matrices** and lift calculation by applying **two sample t-test**. Performed **hyper** **parameter** **tuning** for developing better models and reducing overfitting.
* Implemented **A/B testing** strategies to analyze impact on **KPIs** with understanding of **Project Management** and **Agile** methodologies.
* Created comprehensive dashboards to communicate insights on performance metrics to the stakeholders using **Tableau**. Restricted data for specific users using Row level security and User filters. Created calculated fields, actions, parameters and calculated sets for preparing dashboards and worksheets in **Tableau**.
* Worked with the **Apache** **Spark** Framework for customer analytics using **Spark** **SQL** queries on large scale datasets. Developed flawless **CRM** campaigns and deployed them through multiple channels using **Marketing** **Automation** tools like **IBM Unica**.
* Developed an **ETL** pipeline for loading and mapping data from **SSIS** to **IBM UNICA** for daily **CRM** workflows.

***Analyst at Intelenet Global Services:***  **(*Sep 2015- Feb 2016*)**

* Optimized complex **SQL** scripts for quality checking of projects and populating output tables for **deployment**.
* Tracked **KPIs** and built **Regression** models to predict **customer life-time value**, enhance propensity and scoring attributes.
* Automated hourly status report saving **10 man-hours/week**, thus decreasing response time for fixes and campaign failures.
* Extracted insights and created reports using **Tableau** & **advanced** excel **formulas**, **Excel** **VBA** (Macros), **pivot** **tables** and **slicers**.
* Formulated ad-hoc reports based on requirements gathered from various stake holders using **JIRA**.

**PROJECTS**

***Anti-Money Laundering AML Model (Language/Tools- Python, Jupyter Notebook):***

* Applied models like Logistic Regression and Gradient Boosting to classify the customers as suspicious or not suspicious by data manipulation, feature engineering and variable selection techniques.
* Used various correlation tests and cross validation to reduce overfitting and achieved a recall (sensitivity) of 82% for classifying suspicious customers.

***Topic Modeling on 20-Newsgroups dataset (Language/Tools- Python, Jupyter Notebook):***

* Imported the corpus from the json source using pandas read json command and created bigrams, trigrams using term frequency inverse term frequency (TF-IDF) vectorizer.
* Built Latent Dirichlet Allocation (LDA) model to discover different topics hidden in the corpus and visualized them using the interactive chart in pyLDAvis package.

***Predicting work authorization status (Language/Tools- Pyspark, Spark Sql, Databricks Notebook):***

* Used the Random forest ensemble from pyspark ml library in order to predict lottery selection on the Kaggle dataset (3 million records) with an accuracy of 91% on the Databricks platform.
* Used chi-square correlation test and spark sql to analyze, manipulate, clean and merge data before building a pipeline to pre-process both train, test sets by applying the indexer, assembler and one hot encoder.

**SKILLS**

* ***Programming:*** SQL, Python, R, SAS, Pyspark, HTML, C#, Excel VBA (Macros).
* ***Big Data Ecosystem***: Spark, Hadoop, MapReduce, Hive, Pig, Kafka, Flume.
* ***Cloud Technologies***: AWS (S3, EC2, Lambda, Athena, RDS, Redshift, EMR).
* ***Tools:*** Tableau, Power BI, Azure ML Studio, RStudio, Jupyter Notebook, Visual Studio, SAS Enterprise Miner, SAS CI, IBM-Unica, SSIS, MS Office, Jira
* ***Libraries*:** Numpy, Pandas, Matplotlib, Seaborn, Scikit-Learn, Keras, Nltk, Gensim, Scipy, Beautiful Soup.