# Mukil Raj Kannan

469-600-1046                   mukilraj.kannan@gmail.com                   www.linkedin.com/in/mukilraj-kannan/

## EDUCATION

**The University of Texas at Dallas, Richardson, Texas**                                                                                    December 2020
*M.S., Information Technology and Management*                                                                                                **GPA-3.75**

**Anna University, Chennai, India**                                                                                                                         March 2016
*B.E., Electronics and communication Engineering*                                                                                                  **GPA-3.8**

## TECHNICAL SKILLS

Programming Languages        : Python, R , SAS , Linux Shell Scripting , Java , JavaScript , C++
Data Visualization Tools        : Tableau 2020.2 , PowerBI ,Quicksight , QlikSense , IBM Cognos , Plotly , SSRS
Databases                              : MySQL, Oracle, SQL Server , IBM Netezza, DB2, Teradata, MemSQL, MongoDB, Cassandra , RedShift
Data Science                         : Python (Jupyter, Spyder), R Studio, Regression, Classification, Clustering, A/B Testing, Neural Networks, NLP, Deep Learning (CNN and LSTM models), Regularization, Optimization of models, GANs
Big Data                               : Hadoop, MapReduce, Apache Spark , Hive , HDFS, Spark , Beeline , Pig , Sqoop
Data processing tool (ETL)    : Informatica, IBM Datastage 9.1 & 11.5 , Pentaho , AbInitio, SSIS , Alteryx, BMC Control M ( Job scheduler )
Other technologies                : Git , Jenkins , JIRA , Advanced Excel ( VBA , VLookup and Macros ) , Agile methodologies , RestAPI

## BUSINESS EXPERIENCE

**Katch Media, Dallas, Texas**                                                                                                                July 2020–September 2020
*Data Science Intern*

- Analyzed and scraped **10000+** movie metadata from external sources like Wikipedia, IMDB and IMSDB using packages such as Beautiful Soup, Selenium and Wikipedia API, that could be scalable up to 500,000 movies and build a pipeline to store them in GCP.
- Spearheaded the **Genome project** as POC for helping the genome analyst to automate the movie tag entries by building machine learning models from scratch to detect locations represented in the movie.
- Designed **Script Analysis Dashboard** to analyze movie sentiment arc, character analysis, identify **KPIs** and **KRAs** for movie production clients based on demography and provided feedback , key insights to Marketing and Senior Executive team
- Developed dashboards, reports, and provided research analysis to executives in verbal, visual, and written media.

**Cognizant Technology Solutions-Chennai, India ( Client : RBC )**                                                            June 2016–December 2018
*Data Analyst*

- Designed the **Global Revenue Dashboard** for countries from multiple continents by gathering data from multiple databases and streamlined the incremental load process to generate reports and automated it to be received by the **Executive Panel** on a daily basis
- Decreased cost and manual work by **80%** by automating Status Reporting email using Python and implemented custom Pig scripts to fetch data from MySQL , Oracle databases and load it into Hive
- Analyzed, engineered and optimized frequent data-fetching SQL queries that improved scalability of the product by **25%**
- Implemented Churn Prediction classification algorithms to predict customer **Churn Probability**, this helped the **RBC Marketing team** to identify customers who has high probability of churn and have a marketing strategy accordingly.
- Generated **40+ ad hoc reports**, sub-reports, drill-down reports, drill-through reports and parameterized reports to provide visible data for senior data analysts and business using Tableau
- Created **20+ interactive Tableau Workbooks** indicating 40+ metrics for **Spend Analytics** for individual customer based on ad hoc request
- Awarded **"Pillar of the month"** twice by troubleshooting complex data issues, resolving incidents and returning the application back to service with minimum downtime and documented the issues proactively for fixing future job failure

## ACADEMIC PROJECT

**Machine learning on GPU Kernel dataset: Python**                                                                                    January 2020-April 2020
Performed experiments with hyperparameter tuning to find optimum parameters using linear regression, logistic regression and SVM models with low cost on GPU Kernel dataset and obtained a maximum accuracy of 97.83% with logistic regression model.

**Analysis of truck fleet information: HDFS, Tableau**                                                                                     June 2019-August 2019
Imported truck fleet data into Hadoop using Sqoop and loaded the imported files into Hive by creating tables. Analyzed the high-risk trucks and drivers by connecting Cloudera HDFS to Tableau using JDBC Drivers.

## COMPETITIONS

**VIZIO - Data Visualization - FINALIST ( UT Dallas )**                                                                          August 2019 – September 2019

- Developed Tableau workbooks from multiple data sources using Data Blending and Cross Database join techniques
- Created action filters, parameters and calculated sets for preparing dashboards and worksheets in Tableau and presenting to Sales Team