# VINAY KUMAR SINGH

Vinay.jobs0801@gmail.com | 469-395-2553| linkedin.com/in/vinaysingh08/| https://github.com/vinay99711/

**EDUCATION**

**University of Texas at Dallas**: Master of Data Science and Business Analytics **Aug 2018 -May 2020**
**Army Institute of Technology**: Bachelor of Engineering in Electronics and Telecommunication **Aug 2011 -May 2015**

**TECHNICAL SKILLS**

**Programming:** Python, SQL, R, Tensorflow, PyTorch , SAS, PostgreSQL, MySQL, Oracle, MS SQL, Teradata
**Big Data:** Hive, Hadoop, Apache Spark, Apache Kafka, MongoDB, Teradata, Spark ML, Spark SQL
**Business Intelligence:** Tableau, R (ggplot2), Python (Matplotlib, Seaborn, Plotly, bokeh, interactive), MS Excel
**Machine Learning:** Clustering, Classification, Time Series, Regression, CNN, Deep Learning and Predictive Analytics

**PROFESSIONAL EXPERIENCE**

**Advanced Analytics Intern, Hallmark Cards, Kansas City, USA** **May 2019 - Aug 2019**

- Designed metrics and scraped ~1M tweets data for **Forecasting** seasonal cards sales (Python, Spark Streaming)
- Defined/Designed data pipeline and data integration to collect, clean and store large scale, cross functional dataset
- Applied advanced Statistical techniques and Mathematical analysis to understand Seasonal card sales **Time Series**
- Performed ARIMA modeling development, validation, implementation and experimentation to predict card sales
- Deployed **Tableau** dashboard sourcing data from the data lake to detect insights and provide actionable recommendation
- Reduced the forecast error rate by 70% which lead to reduction in wastage of inventory assets and saved ~$ 1M

**Data Scientist, Mu Sigma**

**Health Care Analytics, Bangalore, India** **July 2016 - July 2018**

- Investigated and evaluated Machine Learning and Deep Learning Algorithms on multiple Drug sales to perform **Predictive modelling** the Sales of drugs using RNNs, CNN, LSTM on TensorFlow using Time Series data
- Experience in working on Unstructured and Structured database to perform ETL operation of **Data integration and migration**
- Instrumental in researching, prototyping, designing, implementing and evaluating machine learning models using Prescriber Sales rep feedback data for **Multi label Classification** of the performance of Sales Representative
- Digging deep into the millions of data records using distributed computing (Hive/Hadoop) and performing quantitative analysis using Jupyter notebooks with Pandas/NumPy/matplotlib and communicating **Text Analytics** using **Data Visualization**
- Researched prototype, built features and optimized the state-of-the-art Machine Learning and Deep Learning techniques like SVM, Logistic Regression, Naïve Bayes, LSTM, CNN, RNN in TensorFlow in CPU/GPU environment
- Applied various transfer-learning techniques using pre-trained word-embedding like Glove, Word2vec for text similarity
- Effective targeting of Sales rep to prescriber lead to increase drug penetration by ~30%

**Retail Analytics, Bangalore, India** **June 2015 - July 2016**

- Worked with online email direct marketing team to optimize cost and increase customer penetration (Hive, Apache spark)
- Designed **Customer segmentation** model metrics and created large database through data integration and ETL pipeline
- Demonstrated Feature Scaling, Dimensionality reduction and feature transformation using PCA and correlation plot
- Performed **Exploratory Data Analysis** to analyze data by doing Hypothesis testing using Statistical and mathematical models
- Used unsupervised algorithms like K-means to segment customer into cluster for effective targeting
- Provide insights and actionable recommendation in an Agile environment and identify customer to target
- Increased the Return on Investment, **Customer acquisition** by 40% & 14.5% and decreased customer churn rate by ~10%

**Research Student** **Aug 2018-June 2020**

- **Novel COVID 19 Detection Model using Computer Vision:** Performed Data Augmentation technique to increase COVID 19 chest X-Ray Image dataset by a factor of 10 and developed COVID Chest X-Ray Image Classifier based on Classical **Deep Transfer Learning** technique to achieve an accuracy of 85%
- **Deep Learning Sequence Modelling for Sunspot Prediction:** Design Time series model for predicting the number of Sunspot using Sequence Models on Classical Deep Learning framework achieving a MAE of 8.98 (LSTM, GRU Recurrent Neural Network
- **Real Time Tweet Sentiment of Sports Team using Kafka Streaming:** Developed Real time end-to-end sentiment tracker of 10+ sports teams after scraping more than ~1 million tweets helping teams and fans understand overall public opinion