Sai Prasanth Thota
Ph : +1 617-372-4874 | saiprasanththota15@gmail.com | https://www.linkedin.com/in/sai-prasanth-thota/

Technical Skills

Databases & Cloud	: MySQL, SQL Server, MongoDB, Oracle, PostgreSQL, Azure, AWS, Elasticsearch.
Programming Languages	: MATLAB, R, Python, SQL.
Data Integration	: Talend, SQL Server Integration Services (SSIS), Alteryx.
Business Intelligence	: Tableau, Power BI, Qlik Sense, Qlik View, Advanced Excel.
Tools	: Excel, Jupiter Notebook, Toad Modeler, NetBeans IDE, VS code, ER Studio.
BigData Tech	: Apache Spark, Hadoop, Kafka

- Expertise in Machine learning and Deep learning Algorithms
- Expertise in writing complex SQL queries, Visualization, ETL workflows and defining Pivot tables

Education

Northeastern University, Boston, MA   **Jan 2018 – Apr 2020**
Master of Science in Information Systems

Vellore Institute of Technology, Vellore, India   **Apr 2013 – May 2017**
Bachelor of technology in Electronics and Communication.

Experience

Data Analyst Co-op, **Wayfair**   **Jun 2019 - Dec 2019**
- Built **ETL** pipelines for dashboards, validated collected data, and identified the root cause for data discrepancy, improved and corrected **SQL** codes for more accurate performance
- Built SQL queries on large scale data to investigate Ad business on customer segmentation, vertical, countries, evaluated the results of campaigns, forecast ROI for different marketing channel, applied machine learning models to find out potential paid customer, insights for important customers' behaviour, and made recommendations for new campaigns
- Conducted quantitative analysis for user acquisition and monetization through aggregated data, segmentation for targeting audience including SQL/Python to clean up, validate, analyze, visualize data and generate reports
- Conducted A/B testing, improved the experimental design according to the business KPI of the Store marketing team
- Improved features, identified and fixed data gaps on recommendation systems, researched and recommended new data tools and communicated between the business, product manager, data scientist and the engineer team to ensure the data quality

Data Analyst Research Assistant, **Northeastern University**   **Mar 2018 – Jun 2019**
- Performed statistical analysis for deeper understanding of **PUMA** customer demographics; Performed Clustering analysis on Developed automated tool to quantify potential orders; Reduced time through analytical processing from 4 hours to 3 minutes
- Support other teams to answer complex business questions by using ad-hoc SQL queries and visualization

Data Analyst, **Grossmacht Engineers Private Limited**   **Jun 2016 – Jan 2018**
- Extracted data using **SQL** and created interactive dashboards in **Tableau** to help visualize incoming order
- Developed **ETL** workflows using **SSIS** to consolidate costumer data coming from different sources to SQL Server
- Built interactive Tableau dashboards for Management to measure the progress of material supply migration and to improve customer retention and presented to the executive team
- Investigated and conducted study on forecasts, demands, & capital for products; reported directly to the management
- Devised procedures to source data from **APIs** to **Excel** models, automated data processing and transformation using python
- Built web scrapers in **Python** to streamline the data collection from several competitor websites for supporting business needs

Academic Projects

Good Reads Recommendation Engine | Python, Machine Learning   **Jan 2019 – Mar 2019**
- Developed a recommendation engine for books in Python using a Goodreads Dataset from Kaggle
- Compared performance of various models including a Naïve Model using KNN (Cosine and Euclidean si Similarity), Collaborative Filtering based model using SVD and Content-Based Filtering based model using TF-IDF scores

Consumer Insights | *AWS S3, Flask, Spark, Elasticsearch*   **Jan 2019 – Mar 2019**
- Developed a Flask application that allows manufacturers to learn Consumer Insights for product enhancements
- Architected a distributed data pipeline that utilizes *Amazon Customer Reviews* of *size 80GB* and updates indices
- Introduced customized tweaks to **Elasticsearch** indices to improve search relevance on the product reviews

Data Integration and Visualization for a Retail Store | *Talend, SSIS , MySQL, Oracle, SQLServer, SQL*   **Aug 2018 – Dec 2018**
- Developed an enterprise wide data lake that was used for sales analytics and built interactive dashboards in Tableau and Microsoft Power BI that helps in effective Business analysis
- Developed ETL workflows using Talend to confirm data coming from multiple sources into a standardized data lake on SQL server
- Profiled various data sources using SSIS and lookup conditions for appropriate data to load a data warehouse of a company

News Classification and Sentiment Analysis | Python, Docker, AWS, NLP   **Aug 2018 – Dec 2018**
- Web Scraped data from websites; performed EDA to determine most popular new category and author. Performed series of experiments with both CNN & RNN to perform article classification with benchmark considering time and accuracy
- CNN with little hyperparameter tuning and static vectors achieves excellent results on multiple benchmarks; Generated a luigi pipeline using docker and uploaded the results on **AWS S3**; serialized python objects using pickle

House Sales Price Prediction, Northeastern University   **Aug 2018 – Dec 2018**
- Predicting the sale price of houses with 79 variables describing every aspect of the residential houses in Ames, Iowa.
- Efficiently applied Random Forest with Gradient Boosting and XGBoost to predict prices of houses