SUHAS KOTAKI

skotak2@uic.edu | +1-(312)-918-9274 | LinkedIn | GitHub | Chicago, IL

Astute analytics professional with **3+ years** of experience building analytical platforms to accelerate business performance. Adept conceptual understanding of **machine learning** algorithms and **statistical analysis**. Proficient in **SQL, R, Python, Alteryx** and **Tableau** and **cloud platforms**. Vigour to optimize and improve performance of the business systems. Ability to liaison across teams with effective business communication.

EDUCATION

Master of Science in Business Analytics, *University of Illinois at Chicago (UIC)*, (GPA: 3.88/4) **Aug 2019 – May 2021**

Bachelor's in Mechanical Engineering, *Visvesvaraya Technological University, India*, (GPA: 3.84/4) **Aug 2012 – Aug 2016**

CERTIFICATIONS

Business statistics and analysis specialization by ***Rice University*** (**Cert No**: 583FH9LBCHSC)

Python for Data Science with IBM Watson - Cognitive services by ***IBM Certification*** **(Cert No:** XN2KZMBSXDK4)

TECHNICAL SKILLS

Software Tools **:** R, Python (NumPy, Pandas, Scikit-learn, PySpark, PyTorch), SQL, Git, Tableau, Alteryx, Stata, Excel, VBA
Analytics **:** Hypothesis testing, Regression (Linear, Logistic, Ridge, Lasso), KNN, Association rules mining, PCA, Ensemble Modelling
Models Deployment: AWS (EC2, Lambda, ECS), Microsoft Azure ML, Heroku, R Shiny, TabPy server
Coursework **:** Statistics, Data Mining, Healthcare Analytics, Big Data, Operations Management, Deep Learning and Data Visualization

PROFESSIONAL EXPERIENCE

Onco Care Analytics, Chicago **Jul 2020 – Present**

Data Science Intern, *(Python, R, Alteryx, Tableau, Excel)*

- Spearheaded the analysis of **Insurance claims** based on oncology Care Model, identified potential **cost saving opportunities** for healthcare providers to gain the benefit of incentives worth **$3.6 million.** Built end-to-end data pipeline using Alteryx to build insightful visualizations
- Published interactive Tableau dashboards, derived revenue leakages based on cost factors such as drugs, comorbidities, and readmissions
- Devised a linear model based on frequency of supportive and delivery procedure and estimated the cost of radiation therapy
- Designed a predictive model using **Cox Regression**, improvised triage assignments and reduced readmissions worth **$30k+ per patient**

University of Illinois, Chicago **Mar 2020 – Jun 2020**

Research Assistant, *(Azure, STATA, R, Python, VBA)*

- Pioneered the research concept to perform "**Temporal Analysis of Linguistic Depression Markers"** using machine learning techniques
- Extracted facial characteristics of profile picture using Azure Face API, developed the data pipeline to **automate 240k+ tweets extraction**
- Orchestrated topic modelling based on **LDA** and **anchored - CoRex** algorithms, analysed the distribution of topic markers across the timeline
- Implemented case-crossover study design and quantified the occurrence of linguistic markers along pre-post diagnosis phase using conditional logistic regression. Determined, "**the presence of depression markers decease by 30%"** in post diagnosis period

Tata Consultancy Services Limited, Bangalore **Feb 2018 – Aug 2019**

Systems Engineer (*Python, Dashboarding, SSIS, Aspect Dialer, List Management system*)

- Collaborated with the Executive management of American Express, enhanced and maintained the customer contact management platform
- Improved the performance of the existing ETL batch processes and SQL queries and automated the report generations across 5 global markets
- Published real time reporting dashboards for monitoring of production systems and **increased the platform availability time by 8%**
- Implemented automated purging mechanism and reduced the manual efforts, saving **$1.3 million** of production maintenance costs.

Asst. Systems Engineer, *(DataStage, Cognos Analytics and Business Intelligence)* **Feb 2017 – Jan 2018**

- Performed customer cohort analysis of Telecom Products, analysed **customer retention,** built end-to-end ETL pipeline for data visualization
- Identified geographic locations with high customer retention rate, accounting for **$1.6 billion** sales, using Cognos analytical dashboards

REAL TIME APPLICATIONS

Context based Product Recommendation System – Amazon Dataset (*PySpark, NLTK, AWS-ECS, Heroku, Docker, Flask-API***)** **Oct 2020**

- Designed a **big data pipeline to process 1 million records** of user reviews for groceries, incorporated text cleansing tasks and K-means clustering model in the pipeline. Developed a real time application for the recommendation system. Deployed the model on cloud.

PROJECTS

Identify factors influencing users to access Electronic Medical records (*R, Excel***)** **Jan 2020**

- Designed the conceptual model based on domain expertise, extracted data from Health Information National Trends Survey (HINTS) database. Performed extensive feature engineering using PCA and statistical tests, cleansed the data for developing predictive model
- Identified the factors - **age group, technology self-efficacy and income group** to be the influencing variables for EMR usage.

Image Captioning with visual attention – Flickr Dataset (*VGGNet, Keras, GloVe***)** **Nov 2020**

- Developed CNN based encoder and GRU based decoder model for generating captions of an image. Incorporated local attention mechanism to improve the quality of captions. Incorporated pretrained model – VGGNet for CNN based encoder network.
- Performed evaluation against LSTM based decoder model with global attention mechanism, analysed the metrics using BLEU 3,4 scores.

Predicting Parkinson's disease through speech characteristics *(R, R-Shiny and Tableau)* **Apr 2020**

- Identified the disease characteristics by interacting with domain experts. Performed exploratory data analysis of the clinical dataset and incorporated statistical tests, KL Divergence technique to **identify 7 speech characteristics** as predictor variables.
- Built SVM classifier with **84% accuracy,** developed interactive user interface (UI) using R shiny for performing predictions **May 2020**