ABHILASH HOSAAGRAHARA NAGARAJA

+1 (408) 341-5201 | Chicago, IL | abhilashhn1993@gmail.com | linkedin.com/in/abhilashnagaraja/ | abhilashhn1993.github.io/

SUMMARY

Analytics professional with strong quantitative skills and **3+ years** of experience in programming and analytics. Highly skilled in **Python, SQL, R,** and **Tableau**, and adept at implementing **Machine Learning solutions** to enforce key business decisions. An effective communicator, result oriented with a proven track record of excellent leadership, time-management, and presentation skills.

EDUCATION

Master of Science in **Management Information Systems (Data Science & Analytics),** *University of Illinois at Chicago* (GPA 3.77) Aug 2019 – May 2021

Bachelor of Engineering in ***Information Science & Engineering*,** *Visvesvaraya Technological University, India* (GPA 7.5/10) Aug 2012 – Jun 2016

SKILLS

Technical: Python (pandas, numpy, scikit-learn, matplotlib), R, SQL, Java, MapReduce, Spark, Tableau, Excel, SPSS, Git
Database & Big Data: MySQL, DB2, Hadoop, HDFS, Hive, Spark
Analytics: Inferential Statistics, Predictive Modeling, Hypothesis Testing, Parameter Tuning, Text Analytics, Recommendation systems
Courses: Statistics, Machine Learning, Data Mining, Healthcare Analytics, Advanced Database, Big Data Analytics, Information Systems, Marketing

PROFESSIONAL EXPERIENCE

DATA SCIENCE INTERN at Onco Care Analytics LLC **|** *Python, Tableau, DataRobot, claims analytics* Oct 2020 - present

- Orchestrating statistical analysis on **CMS Medicare claims** data to deliver analytical insights in the **value-based-care Oncology Care** model
- Engineer insightful solutions to practices via interactive **Tableau dashboards** to reflect expenditures across **HCPCS/CPT codes**
- Formulate a blend of predictive models to **forecast and report drug expenditures** using **EMR** data for various OCM practices
- **Identified revenue leakages** in expenditures across 10+ key metrics such as **readmissions, hospice, ER visits, regimens**, etc.

RESEARCH ASSISTANT at University of Illinois **|** *Healthcare Analytics, Topic Modeling, NLP, LIWC* May 2020 – Sep 2020

- Spearheaded extensive research aimed at deploying analytical solutions to assess twitter feed of patients with diagnosed mental conditions
- Identified **230** self-declared depressed users on Twitter, extracted **120K+** tweets, pre-processed using **stemming and lemmatization**
- Programmed **LDA** and **Anchor word Corex Topic modeling** techniques to study the progression of Mental & Physical symptoms of the users, and recognize patterns in the linguistic characteristics using LIWC in the pre & post-diagnosis period of depression

DATA ANALYST at Manhattan Associates, Bangalore | *Java, SQL, Supply Chain, Project Management* July 2016 – Jun 2019

- Instrumental analysis in identifying potential pain points in the Transportation Lifecycle Management system and device technical solutions using Java, PL/SQL, XHTML to substantially **improve the performance of the customer operations**
- Designed PL/SQL scripts to enhance the surge handling capacity of the **Shipment planning engine** to accommodate **12%** more orders
- Delivered **20+** technical resolutions aimed at improving the efficiency of Trip planning algorithm thus reducing trip costs by **~15%**

PROJECTS

Racial bias identification in Machine Learning setting | *Python, L2-regularization, Hypothesis testing, Gradient Descent*

- Analyzed the data elements of **35k instances** of stop and frisk and conducted **t-tests & chi-sq tests** to determine significance of features
- Devised an **L2-regularized Logistic regression** model with 10-fold Cross Validation trained using **stochastic gradient descent** to predict arrests
- Tuned probability thresholds and weight parameters to achieve **89% accuracy** and **minimized False positives to achieve 85% precision**

Movie Recommendation system | *PySpark, Alternating Least Squares, K-Means clustering*

- Created a data pipeline by implementing Apace Spark context to load and analyze 27 million user ratings of movies
- Implemented a collaborative filtering approach with **Alternating Least Squares method** and optimized the model to achieve an RMSE of 0.81
- Developed a TF-IDF vectorizer of the movie descriptions and created movie recommendations using **K-Means clustering**

Mortality prediction of Prostate Cancer patients | *R, Random Forest, Survival Analysis*

- Conducted exploratory data analysis to spot key predictors determining the survival rate of the patients with prostate cancer
- Applied **Survival analysis using Cox Regression** on key clinical parameters such as change in PSA levels, progression rate of tumor over time and achieved an **improved Recall** of **84%** over the **73%** recall from the Random Forest classifier

Customer Churn Prediction | *Stepwise Logistic Regression, t-tests*

- Inspected the customer usage metrics of an online business service, recognized key predictors using t-tests & stepwise regression
- Modeled Logistic Regression with forward selection and minimized False Positive rates by **20%** over the traditional Logistic model

Detection of Parkinson's disease using voice measures | *R, PCA, Logistic Regression*

- Performed PCA to identify crucial clinical metrics reflecting the audio frequency samples of patients' biomedical voice measurements
- Employed a Logistic Regression model to classify patients and achieved a prediction accuracy of **82%** and a recall of **84%** on test set

Competitive Assessment of JetBlue Airways | *Tableau, MS Excel, LOD expression*

- Explored Bureau of Transportation Statistics airline data of JetBlue airways to highlight routes with higher profits, minimum departure delays
- Built **informative Dashboards with Network graphs, Time series analysis, Density charts** to design data-driven business recommendations

CERTIFICATIONS

- Neural Networks and Deep Learning by *Coursera* Cert No: **JKMEME7UVGK2**
- Python for Data Science and Machine Learning by *Udemy* Cert No: **UC-TL2QYT30**