

ABHILASH HOSAAGRAHARA NAGARAJA

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

SUMMARY

Analytics professional with 3+ years of professional experience in constructing data-driven solutions to deliver meaningful insights. Experienced in processing large scale data in SQL and implementing Machine Learning and analytical solutions in Python, R, and Tableau. Looking for opportunities to implement my analytical and statistical skills to cater insightful solutions to consumers and businesses.

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 claims data to design, deliver analytical solutions to Oncology practices in the value-based care model
- Charting insightful visuals via interactive Tableau dashboard designs to deliver monthly reports of drug and treatment expenses to providers
- Devised an estimation system with **85% accuracy** to forecast the episode costs by analyzing trends in the usage of radiotherapy treatments
- **Identified revenue leakages** in expenditures across 10+ key metrics such as **readmissions, hospice, ER visits, regimens**, etc.

RESEARCH ASSISTANT at *University of Illinois, Chicago* | *Healthcare Analytics, Data wrangling, Topic Modeling*

May 2020 – Sep 2020

- Spearheaded extensive research aimed at deploying analytical solutions to assess twitter feed of patients with diagnosed mental conditions
- Surveyed 20+ research journals to study the behavioral characteristics of mental illnesses commonly exhibited on social media by patients
- Created end to end pipelines to extract, process, analyze twitter feed and constructed topic modeling templates to capture the underlying trends in the text data

ANALYST at *Manhattan Associates, Bangalore* | *SQL, Supply Chain Management, Project Management*

July 2016 – Jun 2019

- Piloted **quantitative analysis** in the TMS setting to determine the viability of performance KPIs like freight cost, trip time, and trucking capacity
- Designed SQL, Python scripts to generate **informative reports** to empower shippers with carrier tendering thus **reducing yearly trip costs by 15%**
- Guided multiple specialized software enhancements in analysis, design, programming, and maintenance phase for shipper and carrier portals
- Introduced **10+ user-friendly flow lines** using Java, XHTML to substantially enhance the user experience of carrier portal of TMS application

PROJECTS

Racial bias identification in Machine Learning setting | *L2-regularization, Hypothesis testing, Naïve-Bayes*

- 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 Apache 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**

Modeling depression markers on Twitter | *Sentiment Analysis, NLP, Anchor word Corex, LIWC*

- Identified 230 self-declared depressed users on Twitter, extracted **120K+ tweets**, pre-processed the text data 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

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

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
- **Python for Data Science and Machine Learning** by Udemy

Cert No: JKMEME7UVGK2

Cert No: UC-TL2QYT30