

ABHILASH HOSAAGRAHARA NAGARAJA

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

SUMMARY

Analytics professional with **3+ years** of industry experience in constructing data-driven solutions to deliver actionable insights. Experienced in processing large-scale data using **SQL** and implementing **Machine Learning** and **analytical solutions** cross platforms in **Python, R, and Tableau**. Looking for challenging opportunities to implement my analytical skills to tell stories with data and deliver value 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

Programming/Tools: Python (Pandas, NumPy, scikit-learn, matplotlib, seaborn), R, SQL, MapReduce, PySpark, Tableau, Alteryx, Excel, SPSS, Git
Big Data & Cloud: Hadoop, HDFS, Hive, Spark, Azure (ML Studio, Storage, Azure Machine Learning SDK) AWS (EC2, S3), DataRobot
Statistics: Inferential Statistics, Predictive Modeling, Hypothesis Testing, Sampling, Regression, Probability, Confidence Intervals
Machine Learning: Classification, Decision Trees, Random Forest, Ensemble methods, Model evaluation, kNN, Recommendation Systems
Courses: Machine Learning, Data Mining, Healthcare Analytics, Advanced Database, Big Data Analytics, Information Systems

PROFESSIONAL EXPERIENCE

DATA SCIENCE INTERN (*Python, Tableau, DataRobot, claims analytics*)

Onco Care Analytics LLC (Bluerock Healthcare IT) – Chicago, IL

Oct 2020 – present

- Orchestrating statistical analysis on CMS claims data to design, deliver informative reports to Oncology practices in the value-based care model
- Devised a prediction system with **85% accuracy** to forecast the episode costs by analyzing trends in the usage of radiotherapy treatments
- Formulated and tuned a blend of classifiers to generate a monthly forecasting report of Super-Utilizers within an episode with a **recall of 89%**
- Designed interactive Tableau dashboards to indicate revenue leakages across 10+ metrics such as **readmissions, hospice, ER visits, regimens**

RESEARCH ASSISTANT (*Healthcare Analytics, Data wrangling, Topic Modeling*)

University of Illinois, Chicago – Chicago, IL

Apr 2020 – Sep 2020

- Spearheaded the research concept of **Temporal analysis of topics in Mental Health Markers** in Twitter posts of diagnosed individuals
- Configured data pipelines to automate the identification and filtering of self-diagnosed users on Twitter, extracted 250k+ tweets of various exhibits of mental health, and created a demographic inference of users with a combination of *m3inference* and *Azure FaceClient API*
- Created graphical representation of topics distribution over 90 days to assess the progression of various symptoms and health characteristics

DATA ANALYST (*SQL, Supply Chain Management, Project Management*)

Manhattan Associates – Bangalore, India

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, conducted **t-tests & chi-sq tests** to establish statistical relation with the predictor
- 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 positive** counts to achieve **85% precision**

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

- Created a big-data processing pipeline by implementing Apache Spark context to load **27 million user** ratings instances from *Movielens* dataset
- 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 text and created content-based movie recommendations using **K-Means clustering**

Modeling depression markers on Twitter | *Sentiment Analysis, NLP, Topic Modeling*

- Identified 230 self-declared depressed users on Twitter, extracted **120K+ tweets**, pre-processed the text data using stemming and lemmatization
- Programmed LDA and Anchored 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

Package Pricing Prediction at Mission Hospital | *R, ANOVA, Linear Regression*

- Explored Mission hospital patient's data to create a comprehensive feature space including clinical parameters and medical history of patients
- Designed a Linear Regression model to predict base price and additional implant costs with an adjusted R-squared value of 0.89

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

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

CERTIFICATIONS

- Neural Networks and Deep Learning** by Coursera
- Python for Data Science and Machine Learning** by Udemy

Cert No: [JKMEME7UVGK2](#)

Cert No: [UC-TL2QYT30](#)