VIPANCHI REDDY KATTHULA

vipanchireddyk@gmail.com | Chicago, IL 60607 | (312)-918-9344
vipanchikatthula.github.io | www.linkedin.com/in/vipanchikatthula | public.tableau.com/profile/vipanchi

SUMMARY

Motivated Data Scientist with 3+ years of professional and teaching experience processing millions of records and delivering actionable insights. Led a team of 3 and successfully completed 7 projects in 90 days. Developed data-driven models, machine learning applications, and delivered analytics solutions for business decisions. Skilled in Python development, R programming, optimizing SQL queries, and Tableau.

EDUCATION

Master of Science in Business Analytics, University of Illinois at Chicago [**GPA 4.0 / 4.0**] Aug 2019 – Dec 2020
Bachelor of Engineering in Electrical and Electronics Engineering, BITS Pilani [**GPA 7.72/10.0**] Aug 2013 - May 2017

SKILLS

- **Technical :** SQL, R, Python(Scikit-Learn, Matplotlib, Pandas), PySpark, PyTorch, Keras, MapReduce, Hive, Alteryx, Tableau, Power BI, Git, STATA, Excel
- **Cloud :** AWS (EC2, S3, Redshift, Sagemaker, Lambda), IAM Role Management, Docker, Kubernetes
- **Analytics :** Inferential statistics, Collaborative Filtering, Cox Regression, A/B Testing, Time Series Forecasting, Recommendation Systems
- **Coursework :** Data Mining, Healthcare Analytics, Big Data, Marketing, Business Data Visualization, Deep Learning, Revenue Management

CERTIFICATIONS

- Tableau Desktop Specialist, A/B Testing – Udacity, Alteryx Advanced Certified

PROFESSIONAL EXPERIENCE

UNIVERSITY OF ILLINOIS AT CHICAGO *- Information and Decision Sciences*

MACHINE LEARNING RESEARCHER | NLP & Survey Analysis May 2020 – Present

- Identified the factors affecting the use of healthcare wearable devices in the US by engaging with domain experts. Found that people with income greater than $75,000 are 2.6 times more likely to by Fitbits or Apple Watches compared to people with income less than $20,000
- Gathered and processed tweets of 2400 hospitals associated with Medicare to analyze the changes in internal policies, systems, and processes because of COVID-19. Researched and summarized the trends & patterns of topics discussed overtime on Twitter

GRADUATE TEACHING ASSISTANT | Statistical Analysis Aug 2019 – May 2020

- Demonstrated Hypothesis testing and descriptive statistical concepts in R programming in 5 lab sessions of 70 students each
- Aided students with Bayesian statistical concepts, techniques for multivariate regression, time series, and statistical process control

UNITEDHEALTH GROUP *– Hyderabad, India*

DECISION SCIENTIST | Retail Dashboards & Healthcare insights Jun 2018 – Jun 2019

- Collaborated closely with clients, gathered requirements, and conducted exploratory data analysis on clinical data. Integrated Ordinal Logistic Regression model (AUC 94%) into user platforms and published Tableau dashboards to track the KPIs
- Delivery date estimator – Deployed classification models to predict the delay in delivery time of new orders which *reduced* the operational expenses in tracking lost orders. Optimized Hive Queries reduced the runtime by 40%
- Improved the efficiency of ETL pipelines from RDBMS to HBase using Sqoop which decreased the delivery time of downstream BI applications by 20%

DATA ENGINEERING ANALYST | ETL Framework and analysis Jun 2017 – Jun 2018

- Streamlined & maintained data warehouse, data lake metadata, and user documentation for internal business customers. Created and managed database objects including tables, indexes, and procedures using SSIS workflows and data pipelines
- Accelerated the operations for filtering and transformation of the raw data by doing a consistent health check of complex SQL queries over time
- **Key Achievement:** Designed interactive Tableau decision dashboards for the compliance team of a major retailer to understand risks associated with their practices. Helped avert potential non-compliance and regulatory risks worth more than $2.0M

PUBLICATION

"*Use of Wearable Healthcare Devices by US adults: Patterns of Use and Key Predictors*" Chandrasekaran R, **Katthula V**, Moustakas E, Journal of Medical Internet Research. 26/07/2020:22443 [Link]

DATA SCIENCE APPLICATIONS

English to Telugu translator: *(AWS, Python, Keras, Flask, Docker)*

- Developed an LSTM model to convert news articles in English to Telugu. Encased the model in Flask application and launched it live on an AWS Linux EC2 instance. Implemented Docker and Container orchestration to simplify app development and distribution to end-users using Amazon Machine Image (AMI)

Instagram Depression Detection: *(Python (NLTK), Pandas, Matplotlib, OpenCV)*

- Extracted text features, HSV and Face-count values from images and text of Instagram posts of individuals suffering from depression.
- Built Un-supervised LDA and Semi-supervised Topic Models using text features. Modeled Support Vector Classifier to predict the probability of depression and improved accuracy from 70% to 94.5% at 89% recall and 92% precision

PROJECTS

Product Recommendation System: *(Python, NLP, PySpark)*

- Created pipeline for data cleaning and TF-IDF vectorization of product reviews for context-based search result suggestions using K-Means clustering. Built Alternating Least Squares model to recommend groceries by analyzing a user's purchase history with RMSE value of 0.9124

Automated Image Captioning: *(PyTorch, DenseNet121, RNN, LSTM)*

- Implemented transfer learning, LSTM techniques to generate captions for images in Flickr 8K dataset and enhanced the model performance considerably using attention mechanism to capture the subject in the images in a better way

Package Pricing at Mission Hospital:*(R, ANOVA, Linear Regression)*

- Predicted the cost of treatment for patients even before the diagnosis by analyzing their health metrics during admission and previous medical conditions. Designed Linear regression model to estimate a base price and additional implant costs to attain an adjusted R-squared value of 0.89

Amazon's Twitter Sentiment Analysis: *(Python, Pandas, NumPy, PySpark)*

- Analyzed sentiment from 400k tweets on amazon's handle using Spark to achieve better NPS Scores for the delivery service. Modeled a RF classifier with 100,000 features to achieve an accuracy of 93.56%, recall of 90.2%.