Sathvik Thogaru
sathvik.srt@gmail.com

Ph: (413) 425-3155
Sathvikthogaru
ThogaruSathvik

Professional Summary

- 2 years of work experience in generating analytical insights to drive business and stakeholder decisions, using a range of analytical tools and technologies such as Python, SQL, R, MySQL, ETL, Tableau, Power BI.
- Involved in all areas of the software development life cycle including requirements gathering, analysis, design, development and testing of the application.
- Solid understanding of the concepts: Data Collection, Data Cleaning, Data Modeling, Data Visualization, Statistics, Hypothesis testing, Model evaluation, Regression Modeling, Classification algorithms and Machine Learning.
- Good at database programming using SQL: joins, procedures, views, triggers, cursors and complex queries.
- Good experience in building visualization dashboards & reports using Tableau, PowerBI, R Shiny.
- Well versed in using the version control system Git. Hands-on experience in CI/CD and workflow automation setup and usage using GitHub Actions.
- Experience in working in an Agile environment. Used Jira to plan, track and manage project tasks and activities.
- Strong analytical and problem solving skills.
- A great team player. Capable of adapting to the changing priorities and meeting the deadlines in a fast-paced environment.
- Takes initiative. Strong communication skills. Passionate about learning and mastering new skills.

Education

University of Massachusetts Amherst — **Amherst, Massachusetts**
MS in Data Analytics and Computational Social Science **(GPA: 3.8/4.0)** — Aug 2021 - Dec 2022
Jawaharlal Nehru Technological University Hyderabad — **Hyderabad, India**
Bachelor of Technology in Computer Science — Aug 2015 - July 2019
Google Cloud Certified Professional Data Engineer(Link) — Nov - 2022

Skills

Programming Languages	Python, SQL, R, C, Java.
Database Technologies & Tools	MySQL, MS Access, SQLite3, ETL.
Framework & Libraries	Flask, TensorFlow, Keras, PyTorch, OpenCV, SciPy, Seaborn, Scikit-learn, NumPy, Pandas.
Data Modeling Tools	Lucidchart, ERDPlus.
Data Visualization Tools	Tableau, Power BI, ggplot2, matplotlib, Google Data Studio.
Version Control Systems	Git, Apache SubVersion.
Methodologies	Agile, Waterfall.
IDEs	RStudio, Visual Studio Code.
OS & Cloud Platforms	Windows, Ubuntu/Linux, AWS, Google Cloud Platform.
Applications	Drupal, Elfsight, Miro, PeopleSoft, Zoho CRM, Jira, MS Excel, MS Powerpoint, MS Word.

Work Experience

Data Analyst — June 2022 — Present
University of Massachusetts Amherst

- Collaborated with department heads to gather requirements on the areas of interest to perform comprehensive analysis on global research engagement, faculty travel, partnership agreements and student exchange.
- Implemented Data Ingestion from multiple sources, thereby reducing data processing time by a significant margin.
- Developed Analytics Module using Open Source tools, and saved the annual license cost for the IPO office.
- Analyzed and interpreted trends/patterns in current global research engagement to help the department heads determine the research offerings for the future.
- Provided insights for decision making on future initiatives, strategies, and resource utilization.

Software Engineer Oct 2021 — Present

University of Massachusetts Amherst

- Collaborated with stakeholders such as professors, technical writers to gather requirements and build the online and multimodal teaching website.
- Translated requirements into actionable tasks by breaking up the work using Agile development practices & Jira.
- Monitored, reviewed and graded 1000+ student coursework submissions.
- Created workflows using GitHub Continuous Integration and Continuous Deployment (CI/CD) to automate the build, test, and deployment of the coursework publishing process.
- Built data sets and dashboards to provide insights on the effectiveness of the multimodal teaching website, based on student participation and interaction.
- Applied changes to the website using actionable insights, measured the results of those changes and communicated them to the stakeholders.

Data Analyst Aug 2019 — Sept 2021

Publishsutra, India

- Analyzed the impact of the current landing page. Made recommendations to change the landing page and other areas to improve the customer conversion rate. Achieved a 15% boost in free trial activation rate.
- Used clustering analysis to identify the inactive students. Provided actionable insights to the marketing team, thereby increasing renewals and revenue by 19% in a 3 month period.
- Identified common complaints amongst new student enrollments and suggested improvements.
- Monitored results of the applied improvements and measured an improved satisfaction rate of 14%

PROJECTS

Social Media Disinformation and Misinformation Analysis on Coronavirus

Technical Environment: Meltwater, Python, R, Google Data Studio

- Analyzed media content of 20k+ "messengers" 80k+ "info-volunteers" on social media platforms, through Social Listening & Network Analysis. Formulated ‘Good-Influencer Criteria’, considering Post-Sentiments, Geographic-Reach and Advertising Value Equivalency. Devised ‘Data Driven Go-to-Market Strategy’ to hire influencers, for Region/Language specific Social Media Campaigns

Movie Recommendation Engine and Feedback Analysis

Technical Environment: Python, numpy, pandas

- Led a team of 4 and developed a collaborative filtering-based recommendation engine using Matrix factorization in Python. Performed Comparative Analysis between KNN Matrix Factorization using SVD. Analyzed the temporal shift measure in user preferences and recommendations using Monte Carlo Simulation. Findings: Popularity bias in Movie genres: comedy, action, and crime; User preference shifted over time with more movie views.

Multinomial Naive Bayes for Document Classification

Technical Environment: Python, numpy, pandas

- Implemented Multinomial Naive Bayes algorithm to classify movie reviews. Prepared IMDB Large Movie Review Data in Bag-of-Words format for Algorithm Training. Classified instances by Laplace Smoothing managing zero probability scenarios. Assessed algorithm Performance for different Laplace smoothing values, achieving 76% accuracy.

Loan Eligibility Prediction

Technical Environment: Python, numpy, pandas

- Implemented Random Forest Algorithm to predict loan eligibility, based on 8 categorical and 4 numerical attributes. Applied Stratified Cross-Validation technique, assessing algorithm’s generalization potential. Evaluated algorithm’s performance on hyperparameter: ntree, achieving accuracy 83.4%. .