SIDDHESHWARI BANKAR

Seattle, WA • 206-712-5810 • siddheshwaribankar24@gmail.com • linkedin.com/in/siddheshwari-bankar/
github.com/Siddheshwari19 • siddheshwaribankar.wixsite.com/siddheshwari-bankar

SUMMARY

A recent Computer Science graduate with strong foundation in Data Science and Machine Learning. Bringing problem-solving abilities and collaborative research experience as Graduate Research Assistant, highly motivated and eager to apply knowledge gained through coursework, internships, and research in professional setting with a strong desire to enhance skills. Proficient in Python, SQL, Data Analysis, and Machine Learning and actively seeking full-time roles to contribute towards innovative solutions.

EDUCATION

Master in Computer Science, Data Science | *Seattle University, Seattle, WA, USA* | GPA 3.7 **March 2024**
Relevant courses: Distributed Systems, Machine Learning, Data Science, Visual Analytics, Big Data Analytics
Involvement: SU ACM-W club, SU AnIMaL (AI/ML) club, Event Operations Crew
Affiliations: AnitaB.org, Grace Hopper Celebration scholar, Women who code, Rewrite the code, Women in Tech

Bachelor in Computer Engineering | *University of Pune, India* | GPA 3.55 **May 2022**
Relevant courses: Object Oriented Programming, Database Management Systems, Project Management, Data Mining
Involvement: National Service scheme, Event coordinator

SKILLS

Python, SQL, Power BI, Tableau, Linux, Agile, Excel, JavaScript, TensorFlow, R, Hadoop, AWS, Docker, Git, GCP, Spark

EXPERIENCE

Graduate Research Assistant (Data Science) | Seattle University **Mar. 2023 – Mar. 2024**

- Developed incremental synthetic data generation and evaluation system (SDGnE) to address class imbalance problem in health risk prediction, especially for starved and rare events preserving data quality.
- Experimented various ways like incremental boosting with control coefficient on different SMOTE variants to enhance the quality and diversity of generated data, resulting in a 19% improvement in model performance.
- Leveraged transfer learning to train machine learning models and evaluated performance of models using various metrics.
- Released Python package complete with detailed documentation and demo video, providing user-friendly guide to effectively utilize the system, thereby facilitating wider adoption and application in relevant fields.

Data Science Intern | M.B.B. Consulting Engineers and Architects **Mar. 2021 – Sep. 2021**

- Automated data extraction from diverse sources, including project databases and financial records, as part of the ETL (Extract, Transform, Load) process to gather relevant information for analysis.
- Analyzed project expenses to identify cost saving opportunities, saving approximately $300 per project and driving overall business enhancement.
- Implemented process pipelines to reduce data collection and analysis time by 30%, enabling faster decision-making and enhancing overall productivity and business performance.
- Developed data visualization dashboards, empowering stakeholders to make informed, data-driven decisions, leading to a 24% increase in on-time project deliveries.

PROJECTS

Multimodal AI WalkExplorer | Python, Multimodal AI, Excel, GCP, OSM

- Developed a path recommendation system for generating pedestrian-friendly routes based on walkability score using OpenAI's CLIP model, Google Cloud Platform (GCP) and Open Street Map (OSM) python library.
- Evaluated the performance of CLIP model by conducting human survey with over 50 participants, enhancing reliability of perceived walkability score.

US Hospitals Data Visualization | Excel, Tableau, Python, Power BI

- Engineered a data transformation pipeline including data exploration, pre-processing, and outlier detection, utilizing Python for advanced data manipulation and feature engineering.
- Designed and integrated a data model that revealed hidden patterns and correlations, providing actionable insights.

Book Recommendation Search Engine | AWS (DynamoDB, EC2, S3), Flask, Hadoop, Docker

- Programmed and deployed a Flask-based search engine on EC2 that recommends books using text queries to books relevance score and pre-computed TF-IDF (Term Frequency – Inverse Document Frequency) scores from DynamoDB.
- Leveraged Hadoop, MapReduce and AWS for scaling and to enhance data processing efficiency for large-scale text files.

Early Plant Disease Detection | Python, TensorFlow, Transfer Learning, Keras

- Developed an image classification system for early plant disease detection using Transfer Learning and built classification models accomplishing over 90% accuracy.
- Conducted extensive data analysis on a large dataset of various plant images to identify the most suitable plant for accurate disease detection.

PUBLICATIONS

- SDGnE: A Synthetic Data Generation and Evaluation System for Rare Event Prediction, W. D. Bae, S. Alkobaisi, S. Bhuvaji, S. Bankar, accepted and will be published in the proceedings of the 29th International Conference on DASFAA, 2024
- VISITGST: A Visitor Management Android Application (IJIRSET), March 2022, e-ISSN: 2319-8753, p-ISSN: 2347-6710