GOURI BENNI

: msoneai.com : linkedin.com/in/gouri-benni/ | : github.com/gouribenni| : gouri.benni@gmail.com | : (408)-690-4151

Summary

As an enthusiastic Data Science professional, I blend advanced analytics, engineering, and machine learning to transform data into strategic business assets. Skilled in the development and optimization of ETL workflows, SQL, Python, Machine learning, deep learning, NLP techniques, adept at orchestrating sophisticated cloud-based data architectures, and proficient in distilling complex data into actionable insights.

Education

San Jose State University, California, USA Aug 2022 –Present
Master of Science (M.S) in Data Analytics GPA: 3.4

- Coursework: Data Visualization, Data Mining and Analytics, Databases, Stats, Machine Learning, Deep Learning, Data Analytics, Management.

Gogte Institute of Technology, Karnataka, India Aug 2016 – Sep 2020
Bachelor of Engineering (B.E) in Computer Science Engineering GPA: 3.5

- Coursework: Python, Databases Data Science, Data Structures and Algorithms, Software Engineering, Cloud Computing, Artificial Intelligence.

Work Experience

Norfolk Southern (NS) Corp May 2023 – Aug 2023
Data Science Analyst Intern Georgia, USA

- Conducted advanced statistical analysis and A/B testing using Excel and SQL to optimize NS train schedules. Identified key patterns and anomalies, leading to a 25% reduction in train delays, enhancing operational efficiency and customer satisfaction.
- Built the creation of feature stores and data assets for machine learning models by integrating an analytics platform with Delta Lake on AWS. This initiative streamlined algorithmic workflows and accelerated data preparation, substantially reducing time spent by data scientists on model readiness.
- Created Tableau dashboards for real-time monitoring of NS train delays, focusing on departure, arrival times, and delay causes. Improved operational efficiency and reliability, leading to a 15% increase in on-time train arrivals.
- Applied NLP techniques for textual analysis of unstructured NS operational data. Used sentiment analysis, topic modeling, and NER to extract insights, contributing to a 10% improvement in operational decision-making accuracy.
- Utilized deep learning models, including LLM and transformers, for inspecting NS trains and improved anomaly detection accuracy by 30%, enhancing safety and maintenance efficiency.
- Implemented XGBoost and ARIMA models for predictive maintenance of NS rail tracks. Reduced maintenance-related delays by 18% and enhanced lifespan of rail infrastructure.

ADP (Automatic Data Processing) Sep 2020 – Jun 2022
Data Analyst Hyderabad, India

- Developed SQL and Python-based data models in ADP Data-Cloud to address payroll issues like overtime miscalculations and tax discrepancies, enhancing processing accuracy by 25% and cut payroll errors by 15%.
- Built a Data Lake enhancing ADP Assist's AI, boosting predictive analytics for employee turnover and compliance. Achieved 30% faster data retrieval and 20% better processing efficiency.
- Created Python-based ETL pipelines enhancing real-time data integration for ADP Workforce Now. This improvement enabled more dynamic tracking of employee time and attendance, reducing data latency by 35% and ensuring timelier and more accurate payroll and HR reporting.
- Implemented data models and Tableau reports in ADP Vantage HCM, improving HR decision-making by 40%. Provided insights for employee development and talent management strategies.
- Upgraded ADP's data warehouse, enhancing data scalability and query performance. Essential for complex reporting for multinational clients.
- Applied clustering algorithms to segment ADP's client base in ADP Analytics, leading to more tailored payroll and HR services. This approach helped in identifying unique client patterns, increasing satisfaction by 20% and boosting retention by 15% through more personalized service offerings.
- Streamlined project management using Jira for ADP implementation projects, aligning them with business goals by 30%. Continuously learned and applied emerging technologies in AI and ML, improving team productivity and innovation by 20%.

Projects

Gen AI Powered Job Transition Pathway

- Spearheaded a data science project to create an AI-powered job transition pathway using generative LSTM models, synthesizing data from multiple sources including Indeed for jobs, Udemy for courses, and EMSI for skills.
- Conducted thorough data cleaning and processing, employing pre-built NLP tools like SpaCy for data normalization, and developed a fuzzy matching algorithm to associate job progression stages with corresponding skill sets.
- Leveraged BERT embeddings within a Seq2Seq LSTM framework to capture contextual nuances in job descriptions, ensuring the model's deep linguistic comprehension for accurate job and skill predictions.
- Designed and evaluated an Seq2seq model with encoder and decoder that predicts individualized career progression paths and provides tailored course recommendations, utilizing exploratory data analysis (EDA) for synthetic data generation to enhance model training.
- Validated the model using a separate dataset to ensure generalizability, achieving a BLEU score of 51.43% for job predictions and 68.84% for skill predictions, reflecting the model's adept training assimilation and its proficiency in predicting career progression pathways.

Ranked Stack Overflow: Mathematics and Statistical Analysis

- Conducted comprehensive NLP analysis on over 370,000 Stack Overflow threads, developing an advanced ranking model. This model significantly boosted answer retrieval accuracy by 64%, showcasing expertise in applying NLP for real-world data challenges.
- Implemented machine learning classifiers like Random Forest and SVM for content tagging, achieving 96% precision. Enhanced the relevance of solutions through the deployment of 'all-MiniLM-L6-v2' and 'all-mpnet-base-v2' transformer models, reflecting a deep understanding of AI technologies in semantic analysis.
- Pioneered a personalized ranking algorithm, improving NDCG scores by 21%. This innovation, combined with a 30% reduction in query resolution time, significantly enhanced user experience on educational forums, demonstrating a balanced focus on technical and user-centric design.

Skills

- **Programming Languages:** Python (Pandas, NumPy, Matplotlib, Seaborn, SciPy), R, C/C++, HTML, CSS
- **Databases:** MySQL, NoSQL, ETL, MS SQL Server, Nested Queries
- **Tools:** Advanced Excel, Tableau, Jupyter Notebook, GIT, Airflow, AWS, Google Data Studio, PowerBI
- **Machine Learning:** Clustering, Business Prediction, Timeseries Forecasting, Classification and Regression.
- **Project Management tools:** JIRA, Agile (Scrum) methodologies, Gantt chart.