Gouthum

9071101929 | gouthumkharvi1899@gmail.com | linkedin.com/in/Gouthum-Kharvi | github.com/GouthumKharvi

Education

Bhandarkars Arts and Science College — Kundapura, Karnataka
Bachelor of Computer Application — *2016 – 2019*
– Relevant Coursework: C, C++, Java, Vb.net, Asp.net, Excel, SQl, MySql, Python, NoSql, R, PowerBi, Tableu, ML,DL, AI, NLP

Professional Experience

AI/ML Intern — Jul 2025 - Present
Pinaccle labs — *Bangalore, Karnataka*
– Developed AI-powered Resume Parser for automated data extraction and semantic job matching using Sentence-BERT (all-MiniLM-L6-v2) with cosine similarity, building ML pipelines for preprocessing, embeddings generation, similarity computation, and evaluation with scikit-learn metrics. Implemented AI Global Translator with multilingual translation, speech recognition, text-to-speech, and OCR capabilities, enabling multi-format document processing (PDF, DOCX, images) using PyMuPDF, python-docx, and PyTesseract. Integrated Google neural translation APIs, deep-translator, SpeechRecognition, and gTTS for real-time translation, and created interactive Streamlit web apps for instant resume analysis, job recommendations, and translation outputs.

DataScience and Analytics Intern — Jul 2025 - Aug 2025
Future Interns — *Bangalore, Karnataka*
– Worked on projects like social media campaign tracking, e-commerce sales dashboards, and student feedback analysis during internship at Future Interns. Cleaned and transformed data using Power Query in Power BI, built Power BI dashboards for key metrics, and applied NLP for sentiment analysis. Also developed a machine learning model for campaign prediction and deployed a Streamlit web app for interactive reporting.

Technical Support Engineer — Nov 2021 - Oct 2023
Glowtouch Technologies — *Mangalore, Karnataka*
– Delivered specialized support via chat, resolving 7,000+ website, hosting, and email issues while managing 4 chats concurrently with a CSAT over 4.7. Performed diagnostics and troubleshooting for FTP, DNS, and SSL to ensure system availability and performance. Maintained an average handle time below 1,250 seconds through efficient issue resolution and technical accuracy.

Epub Operator — Jun 2020 - Sep 2021
Bit9 Business Solutions Pvt Ltd — *Kundapura, Karnataka*
– Converted and formatted 550+ documents into EPUB files with accurate layout, standard compliance, and thorough quality checks. Embedded metadata to improve EPUB indexing, searchability, and cataloging efficiency.

Claims Adjudicator — Jul 2019 - Apr 2020
Hinduja Global Solutions — *Bangalore, Karnataka*
– Evaluated and processed 175+ claims, managed appeals, and ensured compliance with guidelines for accurate adjudication. Prepared reports, verified invoices, and maintained document integrity to ensure 94% accuracy.

Technical Skills

- **Data Analysis and Visualization:** MS Excel, Advanced Excel, Power BI, Tableau
- **Database Management:** MySQL, NoSQL, MS SQL Server
- **Programming for Data Science:** Python (Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, PyTorch)
- **Statistics and Probability:** Descriptive statistics, hypothesis testing, probability distributions
- **Data Preprocessing:** Data cleaning, normalization, transformation, feature scaling
- **Machine Learning:** Supervised and unsupervised learning, model evaluation, feature engineering
- **Data Engineering:** Data pipelines, ETL processes, data integration, AWS(Sagemaker)
- **Natural Language Processing (NLP):** Text preprocessing, sentiment analysis, classification, topic modeling
- **Deep Learning and AI:** TensorFlow, PyTorch, CNN, RNN, BERT, Hugging Face, Generative AI, predictive modeling

Soft Skills

- Critical Thinking, Continuous Learning, Adaptability, Time Management

Certifications

- Certified in Data Science from UpGrad
- UpGrad Data Science Hackathon Certificate (practical ML,DL skills)
- Certified in AI for All: From Basics to GenAI Practice – NVIDIA
- Certified in GenAI Powered Data analytics job simulation – TATA

Projects

AI-Based Product Defect Detection with Computer Vision, CNN, and Streamlit | *Independent Project*

Built an AI system using computer vision and CNN models to detect defects in product images, integrating a Streamlit web app for easy defect image checking. Created processes to clean and prepare images so the model could accurately find issues. Automated the inspection workflow, reducing manual work and errors. This improved the speed and efficiency of quality checks in manufacturing.

Generative AI Project: TinyTales - Children's Story Creator with GPT-2 | *Independent Project*

Created a GenAI project called TinyTales that generates children's stories from a simple starting prompt using the GPT-2 model. The project runs offline in a Jupyter Notebook using PyTorch and Hugging Face Transformers, without relying on any external APIs. Experimented with prompt design and tuned generation parameters like max length, temperature, top k, and top p to improve storytelling quality. Gained hands-on experience with transformer models and generative text techniques.

Credit Card Fraud Detection - With Machine learning | *Capstone Project*

Developed a credit card fraud detection system using machine learning, focusing on classifying fraudulent transactions from a highly imbalanced dataset.Applied data cleaning, SMOTE balancing, and feature engineering to improve model robustness. Evaluated multiple models (Logistic Regression, XGBoost, Random Forest, Decision Tree) and selected Logistic Regression for its high ROC-AUC (0.97 test) and interpretability.

Ecommerce Product Categorization With Machine learning and Deep learning | *Hackathon Project*

The project aims to build a robust multi-class classification system for accurately predicting product categories on e-commerce platforms. By leveraging Machine Learning and Deep Learning techniques, the model enhances product searchability, inventory management, and personalized recommendations. This improves customer experience, operational efficiency, and strategic decision-making, providing a competitive edge for e-commerce businesses.

Duplicate Question Pair Detection - With Machine learning and NLP | *Hackathon Project*

Developed an NLP solution for duplicate question pair detection using Bag of Words (BoW) vectorization on a large dataset with binary similarity labels. Engineered seven custom features, including question length, word count, common words ratio, and word sharing metrics, to complement the BoW representation, resulting in 6,007 total features (6,000 BoW + 7 engineered). Trained a RandomForestClassifier on the combined feature set, achieving effective semantic similarity detection. Built a robust preprocessing and feature engineering pipeline to convert raw text into numerical features suitable for machine learning algorithms. Applied model persistence using pickle to store the trained RandomForestClassifier and BoW vectorizer. Enabled efficient future inference and evaluation through the saved models and preprocessing components.