Sampreeth Kabadi

samkabadi10@gmail.com +91-9148864710 sampreethkabadi

Education

Acharya Institute of Technology Aug 2018 – Aug 2022
Bachelor of Engineering in Computer Science - CGPA: 8.31/10 *Bangalore, India*

Relevant Coursework

- Data Structures and Algorithms
- Design and Analysis of Algorithms
- Big Data Analytics
- Database Management Systems
- Artificial Intelligence and Machine Learning
- Computer Networks
- Operating Systems
- Advanced Java

Experience

High Peak Software *Bangalore, India*
Data Scientist Aug 2022 – Present

- Developed a comprehensive video annotator tool for FitMasters, a fitness app, enabling automated extraction of key workout details (session type, exercise name, intensity, duration) and **reducing manual data entry by 70%** through seamless integration with Google Sheets.
- Implemented a Bidirectional Long Short-Term Memory (LSTM) model to generate exercise sessions from collected data, improving prediction accuracy. Applied data validation and conditional formatting techniques to streamline inconsistency rectification.
- **Improved column header classification accuracy to 96%** for SmartAudit, a workflow solution for audits and compliance, by implementing a Stochastic Gradient Descent (SGD) classifier in Python and automating the preprocessing of master schedules used in employee payroll and compliance workflows.
- Proposed and innovated the use of Google Cloud Run functions, optimizing document processing and **reducing processing time by 35%**.
- Developed an application for 44th Street, a US-based law firm, that summarizes legal transcriptions of court proceedings. Implemented data preprocessing and applied Optical Character Recognition (OCR) on complex PDF and HTML files, showcasing technical expertise and problem-solving skills.
- Optimized topic clustering and labeling for legal documents using BERTopic and OpenAI's GPT.
- Extracted and analyzed text from diverse formats (PDFs, images) for Confluence, generating informative word clouds that aid data interpretation and enhance decision-making processes.

Data Science Intern Jul 2021 – Jul 2022

- Designed a Python script using OpenCV for a US-based law firm to execute OCR and accurately separate pages from multi-page PDFs, achieving **100%** accuracy through contour detection and line recognition.
- Automated JIRA data extraction using JIRA Query Language (JQL) and enforced checks to ensure weekly data updates.
- Generated visualizations to identify outliers and unfiltered data, and created a dashboard using Plotly Express for effective data representation.

Focally *Remote*
Python Developer Intern Oct 2020 – Dec 2020

- Partnered with colleagues to build a real-time object detection model using YOLO, an object detection model for triggering various tasks, applying machine learning techniques in a practical setting.

Projects

Solar Power Generation Data Analysis

- Analyzed solar power generation data from two plants in India (Gandikotta and Nasik) to identify faulty equipment and predict future power generation for improved grid management.
- Conducted exploratory data analysis (EDA) using Pandas and visualized daily power and weather parameters with Matplotlib and Seaborn to identify underperforming inverters.
- Predicted daily power yield for the next two days using time series forecasting with SARIMAX, optimized via Auto-ARIMA, and compared results with PROPHET, finding SARIMAX to be more accurate.
- **Awarded Best Data Visualization** at Technotsava, a technology festival at my undergraduate university, for effectively visualizing solar plant performance data and highlighting critical insights for maintenance and performance improvements.

AutoShortsAssemble

- Automated faceless AI YouTube video creation using Python, Selenium, and ClipChamp, streamlining video editing with subtitles, voiceovers, background music, and timestamps.
- Scraped Reddit data using Python (Reddit-scraper and Pandas), automating video uploads to YouTube, significantly reducing manual effort for content creators.

AI School

- Crafted an AI-powered quiz generation system using Streamlit, automating text extraction from PDFs with pdf2image and pytesseract for OCR, and leveraging Google GenAI for customizable quiz questions.
- Integrated features to automatically convert quizzes into Google Forms, streamlining content creation and sharing for teachers.

Skills

Languages: Python, SQL

Databases: MongoDB, MySQL

Libraries: Pandas, NumPy, scikit-learn, Matplotlib, Seaborn, Plotly, OpenCV, pytesseract, YOLO, pygsheets.

Tools/Technologies: Git, Google Cloud Platform, Minio, Postman, NLTK, BERTopic.

Certifications

- **Exploratory Data Analysis for Machine Learning** - *IBM (via Coursera)*, May 2021
- **Machine Learning** - *Standford University (via Coursera)*, July 2020 - October 2020
- **Crash Course on Python** - *Google (via Coursera)*, November 2020
- **Python for Data Science** - *IBM*, February 2020

Leadership & Volunteer Experience

National Service Scheme (NSS) Volunteer

- **School Revitalization Initiative:** Organized efforts to improve rural school environments through art and crafts, motivating underprivileged students to attend school.
- **Plantation Drive:** Led a tree-planting campaign on World Environment Day, promoting environmental awareness and conservation among students.
- **Swachh Bharat Abhiyan:** Conducted cleanliness drives in schools, educating students on hygiene and waste management for healthier learning spaces.
- **Innovative Teaching Program:** Developed and delivered interactive science and math lessons to government school children, enhancing their academic performance.

DreamPath Foundation Team Leader

- Led a team of 15 students to document and promote local career opportunities in rural and semi-urban areas, focusing on skill development and employability based on regional resources and traditional trades.