Mrityunjay Samanta

mrityunjay@msamanta.com | linkedin.com/in/msamanta | github.com/samantamrityunjay | msamanta.com

Experience

Data Scientist April 2022 – Dec 2024
GS Lab (Now Nurealm) *Pune, MH, IN*

Project: Use of Vocal Biomarkers to extract health insights
Developed machine learning models for vocal biomarker platform to track respiratory and mental fitness

- Analyzed over 10 million voice-based datasets using diverse methodologies to ensure data quality for models and identify correlations between voice attributes and health indicators.
- Applied feature engineering techniques to select the 50 most relevant features from over 2000 voice attributes for machine learning models.
- Trained and evaluated machine learning models, achieving 75% sensitivity and specificity across gender, geographic, and linguistic variations.
- Collaborated with the App & Backend team to integrate the model seamlessly into edge devices.
- Delivered weekly project presentations to clients, providing strategic guidance for future direction.

Skills: *SkLearn, Xgboost, Spark, Seaborn, PyTorch, Databricks*

Associate Data Scientist July 2021 – April 2022
AlgoAnalytics *Pune, MH, IN*

Project: Financial Intelligence Platform
Built a SaaS platform on AWS that turns financial news, social media and company annual reports into structured, searchable insights.

- Architected an end-to-end data pipeline on AWS, using EventBridge to schedule ingestion from Twitter, Google News and yfinance into DynamoDB and MySQL-based RDS.
- Developed a semantic search service with FastAPI that returns relevant news for free-text queries in under a second.
- Optimized NER, sentiment analysis and summarization models for news and tweets using spaCy and TensorFlow, reducing cloud resource consumption by 50%.
- Fine-tuned a YOLOv5 model to detect financial tables across varied report formats with 92% accuracy.
- Classified detected tables by type using transformer-based semantic search over table headers, and extracted their data from 300+ page PDFs using Camelot.
- Deployed the models as Dockerized microservices on AWS EC2.

Skills: *Python, PyTorch, TensorFlow, SkLearn, spaCy, Transformers, Haystack, YOLO, OpenCV, FastAPI, Docker, AWS*

Machine Learning Intern Aug 2020 – May 2021
AlgoAnalytics *Pune, MH, IN*

Project: Automatic Assignment of Medical codes
Assignment of WHO recognised codes into given discharge summaries, health and nursing reports.

- Developed a hybrid deep learning model involving CNN/RNN architecture for the assignment of medical ICD codes with F1 -score of 0.63
- Finetuned Google BERT and BioBERT model to increase the accuracy of assignment process with F1-score of 0.66
- Gained expertise in text preprocessing using various libraries such as NLTK and Spacy to extract relevant sections of the medical text from its raw form.
- Worked with MIMIC III dataset and used descriptive data analysis to devise a machine learning application pipeline.
- Implemented Logistic Regression/SVM/KNN to improve upon the baseline Non-AI model with F1 score of 0.58 which is comparable to literature value.

Skills: *Python, NLTK, Spacy, SkLearn, PyTorch, Pandas*

Skills

Languages: Python, C/C++, SQL
Machine Learning Frameworks: SkLearn, Spacy, NLTK, Pytorch, Tensorflow
Python Libraries/Frameworks: Numpy, Pandas, FastAPI, Flask, Streamlit, OpenCV, Matplotlib
DevOps: Git, Docker
Cloud Technology: AWS

Education

Indian Institute of Science Education and Research Pune, MH, India
Integrated Bachelor of Science - Master of Science *Aug. 2016 – June 2021*
Focus: Physical Sciences, Machine Learning
Award: INSPIRE Fellowship by Govt of India