Abhilash Subhash Sanap

623-399-5688 · asanap1@asu.edu · Linkedin

Education

Master of Science in Computer Science Arizona State University, Tempe, AZ — Aug 2022 - Apr 2024
Relevant Coursework: *Natural Language Processing, Statistical Machine Learning* — 4.22/4.0 GPA

B Tech in Computer Science National Institute of Technology Karnataka (NITK), Surathkal — Aug 2015 - May 2019
Relevant Coursework: *Object Oriented Programming, Software Engineering, Databases, Operating Systems* — 7.6/10 GPA

Patents

- **Space-Efficient Storage Of System-Monitoring Data** Reduced the storage requirement for CFT by **50%** with an intelligent use of star schema and storage of differenced values for high range variables. **(Granted May 23)**
- **Predicting Errors using CFT** Proposed Federated learning method for deployment and predicted errors using Convolutional Neural Network trained on heatmaps created from CFT collected from customers. **(Granted Apr 23)**
- **Storage Array Error Mitigation** Reduced the recovery time for a failed system by **40%** by finding change points in Code Flow Tensors (CFT) and highlighting the module that failed. **(Granted Jul 22)**
- **Storage Media Scrubber** Predicted optimal time for data scrubbing of storage drives using a Local Factor Outlier algorithm to ensure most efficient utilization of the drives thereby reducing significant cost. **(Filed Jul 21)**

Work Experience

Data Science Intern, **o9 Solutions Inc** — June 2023 - Present
- Reduced planning time by **30%** by automating the detection of new Limited Time Offers, Level Shifts and Change Points in historical data, Holiday Spikes and Dips, Stockout conditions etc.

NLP Research Assistant to Prof. Runger (Project funded by Intel) — Jan 2023 - May 2023
- Fine-tuned the open-source Large Language Model **Databricks Dolly v2** to generate an extractive summary about specific companies from the meeting notes and deployed it on AWS.
- Built an end-to-end pipeline in Python that ingests meeting notes and generates insights such as frequency of mentioned companies, sentiment about and extractive summary for each mentioned company.

Data Scientist, **o9 Solutions Inc** — Jul 2021 - Jun 2022
- Reduced value leakage in supply chain of a beverage giant by **5%** with accurate forecasts of demand using time series algorithms like ARIMA, TBATS, Croston and libraries like Prophet, Greykite.
- Expedited client onboarding from a week to 2 days by creating Python scripts that perform driver-based recursive forecasting using adaptable functions for data preprocessing, modelling, and generating explainable metrics.

Software Engineer, AI, Powermax, **Dell EMC** — Jul 2019 - Jul 2021
- Designed and developed the backend for Data Visualization, Analytics and Machine Learning webapp in Python.
- Reduced the latency of webapp by **15%** by creating new metadata tables, SQL Query Tuning and DB optimization.
- Optimized Resource Commissioning by building dashboards monitoring DB size, webapp performance, ETL status.
- Built the Extract-Transform-Load(ETL) pipeline in Python and Bash for collecting data from customer machines
- Collected Migration performance data to allow autonomous decisions for Data Replication in storage systems.
- Reduced debugging time taken by PowerMax engineers by developing inline queries that fetch system variables.

Software Intern, **Fidelity Investments** — May 2018 - Jul 2018
- Defined key performance indicators (KPI) using **3**+ hybrid data sources and calculated them with optimized SQL.
- Delivered an interactive gamified leaderboard using Tableau to track KPIs for **12 teams** in the business division.

Projects

Emotion Detection on facial image
- Arrived at the 'fittest' CNN model for Emotion Detection using a variant of Evolutionary algorithm
- Reduced the time taken for searching optimal hyperparameters by mutation and cross-over operators

Graph Classification [Dataset: CORA]
- Used Graph Convolution Network, Graph Attention Network and Graph Sage for multi label node classification.
- Visualised the embeddings from each model using tSNE to judge the separation of clusters.

Relationship Extraction in Biomedical Data tackling the Large Input Large Output (LILO) problem
- Utilized context-aware summarization using GPT-3 to improve the performance of relation extraction on Electronic Health Records. Guided by Dr Chitta Baral. Dataset: 2010 Relations Dataset.

GPT-3 Prompt Engineering [Model: text-davinci-002]
- Performed In-Context Learning on GPT-3 model, configured the hyperparameters like Temperature, Stop Sequences to create prompts for many useful tasks including Financial Investments, Proposal Writing and Bioethics.

Truth Verification (NLI) using BERT [PyTorch |RoBERTa model | Dataset: Stanford NLI]
- Trained BERT model to infer whether general knowledge statements hold true under a premise, F1-score: 87

Skills

Languages: Python, C++, SQL, R,Bash · **Tools**: Tableau, Postgres, MongoDB, Elastic Stack, Git, Docker, PySpark

ML: SAS, Classification, Regression, Clustering, Tensorflow, Keras, Scikit learn, BERT, XGBoost, LightGBM, CatBoost Silverkite, Prophet, HuggingFace, Deepspeed, NLTK, Rasa, SpaCy, AB Testing frameworks

Cloud Technologies: Openstack, AWS S3, DynamoDB, Lambda, SQS