Nuthan Kumar

Senior Machine Learning Engineer

Contact

Address
Bengaluru, KA, 560085

Phone
984 595 6557

E-mail
Djnuthan@icloud.com

GitHub
https://github.com/nuthanku mar1994

Skills

- Python
- Machine Learning
- Deep Learning
- Pyspark
- Apache-Spark
- MySQL
- MongoDB
- Reddis
- Tensorflow
- PyTorch
- Sklearn
- Flask
- Sanic server
- Microsoft PowerBI
- Amazon Web Services(AWS)
- Pandas
- OpenCv
- Matplotlib
- Seaborn
- IBM Watson Studio
- Excel VBA Macros
- Time series forecasting
- NLP
- Rasa SDK
- Power BI

Certifications

- IBM Data Science
- Google Cloud Certified: Professional Machine Learning Engineer
- Google Cloud Certified: Cloud Digital Leader
- Microsoft Certified: Azure Data Fundamentals
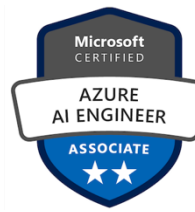
6 years experienced, meticulous & result-oriented Senior Machine Learning Engineer having 4.5 years' experience armed with a proven track record of analytical acumen in deploying complex machine learning models. Experience in analyzing large amounts of user-generated content and process data in large-scale environments using cloud infrastructure. Possesses diverse experience in planning & executing multiple projects and liaising with the key stakeholders to identify & resolve business problem statement and deliver excellent results.

Work History

2022-05 – 2023-08

Software Developer (Machine learning/ Deep Learning)

Gnani Innovation Technologies, Bangalore, Karnataka

BotFramework Platform

- Leveraged Hugging Face Transformer library, spacy for fine tuning Custom NER (payment modes, product names) across multiple Indian Languages and English
- Applied Transfer learning technique using T5 to generate questions and corresponding paraphrases from a given document containing large text.
- Designed and Implemented API layer to integrate bots across various platforms like whatsapp, facebook messenger, slack, microsoft teams and live agent transfer platform Freshchat.
- Implemented Integration layer for various CRMs and Ticketing tools (Freshworks, Zoho and Zendesk)
- Researched, prototyped and developed NLU data analysis tool to identify outliers and anomalies within data and intent overlap using sentence transformers, paraphrase mining and Linear outlier factor Analysis.

Chat/Voice Bots

- Implemented NLU and Dialog Management for 15+ voice bots across various languages (English, Hindi, Bengali, Kannada) using RASA and Dialog Flow.
- Custom pre-processing pipeline to account for ASR inaccuracies in case of voice bots which lead to improved NLU in voice bots.
- Improved NLU accuracy by comparing and feature engineering using multiple word Embeddings (fastText, BERT, m-bert) and SOTA classifier DIET (transformer based intent classifier and entity extractor)
- Scaled the bot to handle 1000 concurrent requests/second and reduced RAM consumption by 80% by designing single model loading across multiple processes, which also led to decrease in deployment cost by 50%.
- Closely worked with DEV-OPS team and Reduced bot deployment time by creating Custom Docker Images and configuration in Kubernetes.
- Reduced Bot Development time and maintainability by implementing common dialog management across multiple languages.
- Implemented topic modelling on custom data containing complaints in various Indian languages to understand the complaint categories in Insurance and Banking domain.

Speech Analytics

- Fine Tuned Cross-lingual Sentiment analysis model (Roberta) to identify verbatim causing objections in code mixed (Hindi-English) text.
- Custom Entity detection using spacy to Detect products and features of a product.
- Prototyped, Generated features and optimized using various machine learning like Linear Regression, SVM, Random forest to classify the user propensity to buy a particular product in a sales call.
- Text-to-Speech System
- Developed Text to Speech System for Kannada and Hindi Languages using Tacotron2 and Wave Glow.
- Multi-GPU Training using Dockers and CUDA 10.1. Optimized Models for CPU inference at 900ms per inference.
- Scaled model serving using custom Docker image and utilizing Kubernetes for deployment.

Retail Scan

- The project is based on object detection and counting of the shelves products in retail chain store in order to automate the process of product auditing of various brands.

Hobbies

- Basketball
- Trekking
- Book reading
- Swimming
- Marathon Running
- Social work volunteering

- Development and implemented M-detectron algorithm in order to train the model and fast deployment onto cameras and hand held devices.
- The algorithm was able to detect, classify and count up to 450 product types with acceptable accuracy.
- Was involved in the complete architecture design of the software.
- **Aadhar photo (Photo occlusion and Face clarity)**
- Development Image classification deep learning architecture to detect any occlusion in face photo and face clarity.
- Achieved **97% accuracy** which beat the second best which could achieve only 84% accuracy.
- Scaled the algorithm to be both light weight and could classify at the rate of 450 images per second.
- Assisted in deployment of the algorithm in cameras and mobile app.

2022-01 – 2022-05	**Financial Data Analyst**

Moody's, Bangalore, Karnataka

- Preprocessed data from 100+ organizational documents and configuring the tenant data to on-board in our platform within 1 hour
- Designed task-oriented chatbot to automate Leave and Request management services to benefit 900+ employees and newly on-board through voice assistant
- Incorporated phrase-boosting along with google voice API to handle domain specific homophonic phrases along with small-talk and abusive language filter
- Extracted all profile related questions by configuring linguistic Dyna JSON API endpoints
- Spearheaded a team of 4 data scientists to create a recommendation system to suggest NSL constructs while building solutions through platform and built orchestrator to integrate the API with UI to be available for end-users

Regression Modelling

- Directed model development, validation, testing, and implementation of analytical products and applications Developed an additive scoring model for QSM & a logistic regression model to yield a K-S statistic of 51.5 Tested and implemented decision trees, random forests, and ensemble models via bagging and boosting
- Dataset Management & Model Integration
- Employed Principle Component Analysis to analyze collinearity and reduce the dimensionality of datasets Applied Bayesian Model Averaging (BMA) to combine individual keyword-level models
- Experience in working with end-to-end **ML Application with Azure MLops Cloud Environment**
-

2021-10 – 2022-01	**Senior Executive**

KPMG, Bangalore, Karnataka

- Conducting economic analysis/ comparability analysis by using Indian and foreign databases. Review of company's balance sheet, profit and loss reports and development the report for the company.
- Development of Power BI Dashboard for the resource utilization, daily work monitoring, Deliverables monitoring.
- Development of functional, assets and risk report for these companies by studying annual report, industry research and client information.
- Worked in **AWS,AWS Glue, AWS Athena** environment for development and deployment of custom Hadoop applications.
- Used Spark and **Spark-SQL** to read the parquet data and create the tables in hive using the Scala API & Play Framework.
- Experienced in using the spark application master to monitor the **spark jobs** and capture the logs for the spark jobs.
- Involved in performing the analytics and visualization for the data from the logs and estimate the error rate and study the probability of future errors using regressing models.
- Experience in building Real-time **Data Pipelines** with **Kafka** Connect and Spark Streaming Using **AWS Lambda Function, AWS Glue, AWS Athena.**
- Involved in developing a **linear regression** model to predict a continuous measurement for improving the observation on wind turbine data developed using spark with Scala API.

2019-05 – 2021-09	**Senior Executive**

Deloitte Haskins & Sells, Bangalore, Karnataka

- Performing data mining for company financial information. Analysis of these information by performing economic analysis and financial analysis.
- Analysing the financial trend of the company using various parameters in order to generate risk analysis of the companies.

- Development of functional, assets and risk report for these companies by studying annual report, industry research and client information.
- Use of MSSQL for data management, data management, and data analysis.
- Performing statistical analysis using machine learning algorithms.
- Producing company performance report by analysing company's financial performance, market conditions, assets owned (tangible and intangibles).
- Create **PySpark** scripts for validating the data which is migrated.
- Created the data frames and worked on transformation
- Understanding the up-stream source nature and work with business cases.
- Develop and monitor script to load the data into redshift tables using **PySpark**.
- data processing and better performance.
- Design the Jobs Using **Pyspark, Python** with **Streaming, SQL Server SSIS migrations etc**.

Programming languages: Python, Java
Machine/Deep Learning: Linear Regression, Logistic Regression, SVM, KNN, Decision Trees, Random Forest, CNN, RNN, LSTM, GRU
Libraries : Pandas, Numpy, scikit-learn, keras, Fasttext, streamlit, matplotlib, hugging face, spacy, nltk, inltk, streamlit
Database : MongoDB, MySQL, Redis, PostgreSQL
NLP : Attention Mechanism, Transformers, BERT, GPT-2, RoBERTA, GPT-3, T5
Cloud: Azure, AWS, Data Bricks, Data Lake.

Education

2017-04 - 2019-04

Master of Business Administration

Christ University - Bangalore

2013-04 - 2017-04

Bachelor of Engineering: *Sir M. Vishveshwaraya Institute Of Technology - Bangalore*