Links:
Linkedin/santhoshkumar
medium/santhoshkumar
GitHub/santhoshkumar
Personal Portfolio

Santhosh Kumar

Skumarr53@gmail.com | +91 953 847 2543

WORK EXPERIENCE

LENDINGKART — **Senior Software Engineer** — **Nov 2022– Nov, 2023**

- Implemented data engineering (**DE**) pipelines for real-time analytics on collection metrics
- Enhanced and integrated new features into the scoring model's rule engine to refine customer loan eligibility verification.
- Responsible for identifying and resolving production issues, ensuring seamless functionality and improved reliability.
- Contributed to seamless integration and continuous development pipeline via **CI/CD** and migrating service to **Kubernetes**, focusing on scalability, automated deployments, and advanced monitoring.
- Contributed to the development of a centralized **Feature Store** to streamline the storage, processing, and access of critical model features.

QUANTIPHI ANALYTICS — **Senior ML Engineer** — **Dec, 2021 – Nov, 2022**

Acord Forms Extraction

- Orchestrated an MLOps-focused comprehensive ML pipeline, facilitating data extraction from Acord forms through advanced AI techniques.
- Leveraged the **LayoutLM**, achieving a Model accuracy of 96% in identifying key-value pairs, checkboxes, and table contents from complex documents.
- Enhanced entity extraction accuracy to 91% via the integration of sophisticated post-processing logic.
- Employed a robust MLOps strategy by containerizing the solution with Docker for deployment on AWS **SageMaker**, ensuring seamless integration into existing workflows and scalable production environments.

Parchment POC

- Developed a solution that extracts institute details, student details, and academic performance from transcripts.
- An accuracy of 94% was achieved by training the LayoutLM model for document understanding.
- Developed a post-process logic to extract semester-wise scorecards from raw model predictions.
- Solution was containerized using **Docker** on SageMaker, seamlessly integrating it into the workflow for optimized MLOps efficiency

Search Engine Development

- Developed a Search Engine that would retrieve relevant documents based on the user search query.
- We experimented with two model approaches, i.e., Probabilistic based (BM25) and Semantic Similarity (BERT)

BRIDGEI2I ANALYTICS — **Analytics Consultant** — **Oct, 2019 – Dec, 2021**

Accounts Payable Fraud Detection

- The objective is to build a machine learning-based Analytics solution that will help constantly flag any fraudulent transactions.
- Used **Azure OCR** to extract data from invoice PDFs, check for certain scenarios, and create meaningful fraud-indicative features around it.
- Derived features from graph networks that help detect fraudulent patterns.
- Designed a Client-facing interactive visualization dashboard using the **Dash** framework.

North America Storm model

- Developed a GLM model to estimate loss cost for North America P&C product
- Daily weather station data was collected from external data sources and tagged to nearby policies to aggregate wind and rain information.
- Relationship between wind speed/rain volume and claims is derived to calculate risk relativity at the zip code level.

VERISK ANALYTICS — **Research Engineer** — **Aug, 2016 – Oct, 2019**

DATA SCIENCE PROJECTS

YOLOv7 Amenity Detection Video Analytics [link]

- Implemented the **YOLOv7** Model to detect Amenities present in the **Video** in real-time.
- Model weights trained on the COCO dataset for detecting object boundaries to benchmark results.
- Fine-tuned model on our custom Amenity data set

Image Caption Generation [link]

- Implemented image caption generation method proposed in **Show, Attend, and Tell** paper using **Fastai** framework to describe the content of images.
- Pre-trained ResNet101 model is used for the Encoder and LSTM for the Decoder. Achieved **24 BLEU** score for Beam search size of 5
- Designed a Flask-based web application enabling real-time caption generation from user-uploaded images.

Animal Pose Detection [link]

- Developed the Animall Pose Detection system trained on a custom dataset of 15,000 images, applying **Detectron2** with the Faster **R-CNN** architecture for accurate pose classification.
- Achieved significant model precision in identifying animal poses with low total loss metrics.
- Designed a Flask-based web application enabling real-time pose detection from user-uploaded images.
- Enabled seamless deployment options, including **Docker** and **Kubernetes**, for scalable application delivery.

Machine Translation - Transformer-based [link]

- Implemented transformer-based **LLM** proposed in **Attention Is All You Need** paper using **Fastai** framework to translate queries French-to-English.
- Achieved **58 BLEU** score using **BERT** architecture.
- Designed a client-facing web application using the Streamlit framework enabling real-time translation.

Amazon Food Reviews [link]

- Built a Sentiment analysis tool that classifies reviews to gain an overview of the customer's opinion in real-time.
- Engineered features from the text using BOW, TFIDF, Word2Vec, Average Word2Vec, Tfidf weighted word2vec featurization techniques
- Optimized k-nearest neighbors, Naive Bayes, SVM, and Logistic classifiers using GridsearchCV to reach the best model.
- **Logistic Regression** with TFIDF Vectorization was the best model with AUROC of **0.982** and **93%** accuracy.

Used Car Price Prediction [link]

- Built an end-to-end feature transformation and model selection pipeline for predicting the price of used Cars, helping buyers make an informed purchase.
- Optimized Linear, Random Forest, GBDT, and XGBoost Regressors using GridsearchCV to reach the best model
- **Gradient Boosting Regressor** turns out to be the best model with **0.033** Mean Squared Logarithmic Error (MSLE).
- Built a client-facing API using **Flask** framework and Hosted on **AWS** using Elastic Beanstalk service via **Docker** image.

EDUCATION

Indian Institute of Science, Bengaluru	M.E in Water Resources Engineering (Civil)	2014 - 16
University of Agricultural Sciences, Raichur	B.E in Agricultural Engineering	2010 - 14

TECHNICAL SKILLS

- **Languages**: Python, HTML, JQuery, SQL
- **Framework**: Flask, Pytorch, Pyspark, OpenCV
- **Tools**: Linux, Docker, Gitlab CI/CD, Kubernetes, AWS Sagemaker, AWS Textract, AWS Lambda, Selenium
- **Skills**: Computer Vision, NLP, Video Analytics, Image Processing, MLops, Cloud services, Generative AI, LLM