Varunsaagar Saravanan

AI/ML Engineer

Phn No : +919944382721
Email id : varunsaagar.s@gmail.com
Linkedin : linkedin.com/in/varunsaagar
Github : github.com/varunsaagar

Working experience of 5 years in Full stack End to End solution on AI/ML with experienced knowledge LLM , Generative AI , Computer Vision, Natural Language Processing (NLP), MLOps, Machine-learning, Deep Learning, Data engineering and delivering as a product and implement action-oriented solutions to complex business.

SUMMARY

- **ML CI/CD Pipeline** - Skilled in Automation of the Data, ML Model & packages encompassing Data standardization, Feature extraction, model validation and optimization with uniformed pipeline and wrap as services that Leveraged to automate daily operational activities.
- **ADAPTION OF RESEARCH PAPERS** – Applying & Using state of art research's into the business requirements and delivering the best AI Architecture.
- **DATA-DRIVEN STRATEGY –**Experienced in using data, analysis and a flare for storytelling to influence the decision-making process at all levels of the organisation towards strategies that deliver value. Discover patterns, formulate and test hypotheses, translate results into strategies for business problems.
- **FEATURISATION & MODELLING -** Collaborate with domain experts and apply their knowledge on data featurisation & machine learning model development delivering more relevant results and conclusions quickly.

AI/ML SKILLS

✓ Advanced Python	✓ Tensorflow, TFlite & JS , Pytorch	✓ EDA & Model preparation, Evaluation Metrics	✓ Deep Knowledge on Object Deduction & Segmentation Models.	✓ AWS, GCP, Databricks
✓ Opencv , Cuda	✓ Airflow, DAG	✓ MLops, Dockering, Kubeflow	✓ WandB, MLflow	✓ Tableau , SQL , Nosql
✓ Versioning control in Gitlab & GitHub	✓ Deployment of POC in Jetson nano, Pi devices	✓ Pyspark , Fast API, Flask	✓ LLM,LSTM, Attention & BERT,Transformer Models,GPT3	✓ Knowledge of statistical methods, Transfer Learning

PROFESSIONAL EXPERIENCE:

Asianet News Media & Entertainment, Bangalore – Lead AI **(June 2022 – Present)**

Roles & Responsibility:

- ❖ Drive the AI/ML vision and strategy in alignment with the company's goals.
- ❖ Implementing LLM and Gen AI product in current CMS and linear Channels.
- ❖ Oversee the design, development, and deployment of machine learning models cycle and Architecting for cloud deployments.
- ❖ Manging vendors for non-AI business initiatives .
- ❖ Heading 5 members of team ,Mentor and guide the AI/ML team, fostering a culture of continuous learning and innovation. Recruit and retain top AI/ML talent.

Projects :

LLM Prompts to AI Anchor Video in regional language.

LLM prompts to generate AI-driven video anchors that deliver news in regional languages. This project aims to revolutionize local news broadcasting.

- ✓ Multimodel GAN with Various Components: with way2lip , RealGan , Bark , SadTalker using pretrained weights .
- ✓ Build UI and warpped with docker and deployed in GCP .
- ✓ Implemented LoRa for LLAMA 2 in the current piplines using internal data to summarize the regional content .

Platforms and frameworks I have worked with :

- *Deployment tools & services* : GCP , Docker , LLMops, Vector DB – Faisis ,Azure cognitive services ,
- *Frameworks/Libraries* : Tensorflow, Transformers , Way2lip , Bark, Lora , RealGAN , mps , Cupy , Cuda
- *Languages* : Python , VD, Gradio, bashscript , ymal

Zee Entertainments, Bangalore – SDE III / AI ML Engineer (May 2022 – June 2023)

Roles & Responsibility

- ❖ Generate Guest users and Uplift the Subscriptions by developing and implementing recommendation models in Zee5 platform.
- ❖ Increase Retention and reduce the churn users by implementing Life time value models .
- ❖ Bringing new AI ML initiatives from Performance marketing team and implement personalised Ads across the social media.
- ❖ Responsible for backend API deployed models and monitoring the metrics for live model.
- ❖ Closely work with business and enhance personalized Cleaver tap notification using ARIMA Models.
- ❖ Creating Tableau Dashboards for Business team and Reco API to metrics daily.
- ❖ Work with cross functionally with Backend , Front end team to flawlessly implement CI/CD pipelines , DAGs , integrate the models on time.

Projects :

Personalized Video Recommendation

The goal of a personalised video ranker is to improve the user's viewing experience by presenting them with content that is more relevant and interesting to them, ultimately increasing engagement and retention.

- ✓ Used Dense and sparse layers with cutting edge activation and custom optimizers which results ***97% Hit ratio*** .
- ✓ Developed a sub models ***content based and collaborative model , auto encoder , Bandit Reinforcement*** to increase the Likelihood Recommendation to power the inputs for PVR model.
- ✓ Created a Fastapi for **inferencing** with a highly optimum low latency inferencing time because model **optimized** with Tensorflow lite which is **5mb**.
- ✓ Made a **Dashboard** for monitoring PVR model **metrics** in **Tableau** which is helpful for business team and stakeholders.
- ✓ Entire Architect is Developed on **MLFlow** dynamic in vertex ai , which is scheduled via **airflow day at daily frequency** .
- ✓ Constructed a data pipeline and standardised a dataset by PySpark and BigQuery for inferencing.

Platforms and frameworks I have worked with :

- *Deployment tools & services* : GCP , Docker, Airflow , Mlflow, Vertexai , Dataproc cluster
- *Frameworks/Libraries* : Tensorflow, TFlite, Fastapi , gunicorn, Deepctr , keras , sklearn , TFR , Cupy , Cuda
- *Languages* : Python, Java , Sql , Pyspark , bashscript , ymal

Impact:

Watch Time and CTR been increased by 67% with this implementation and directly helps in User conversion and engagement .

Creative Thumbnail and Ad generation

An AI intelligent to personalize the video thumbnails automatically for users depending on their taste and cohorts which helps in reducing the manual process.

- ✓ Multilabel Multiclassification has been developed and service runs on demand at metric of 5 percentage increase in effective catalogue.
- ✓ Created an auto thumbnail generative api tool powered by KNN and Opencv which generates 1080p thubmnails .
- ✓ On top of the generated thumbnail genre , blur , person deduction been predicted using transfer learning and computer vison techniques , Models has been developed which also generates the various metadata using TensorFlow .
- ✓ Parallelly developed a user cohort cluster which has taste and other feature. Mapped those AI generated thumbnail to user cohorts which associated with genre level personalization .
- ✓ Entire pipeline is optimised at very low frequency to inference and wrapped as container and served as API .

Platforms and frameworks I have worked with :

- *Deployment tools & services* : GCP , Docker, Airflow , Mlflow, Vertexai ,Vision api
- *Frameworks/Libraries* : Tensorflow, TFlite, Fastapi , gunicorn, opencv , keras , CCP , Cupy , Cuda
- *Languages* : Python, CCP , Sql , Pyspark , bashscript , ymal

Impact:

Impression and Click rate has been increased to 72 percent which helps in User conversion and engagement.

LTV Model and Performance model

Scope of the model to predict the customer life cycle and ad spend value for customer which is decision driven metrics to user level.

- ✓ CLTV estimation, Used regression based ensembled and estimators model , which benefit from the weak learners' ability to cancel out their erroneous estimations.
- ✓ Datapipline been pre-processed , encoded at low float level with high features and own evaluation metric been formulated.
- ✓ Using the predicted cluster the personalised ad will be been rolled by performance marketing team and same data will be used for PVR models .
- ✓ Developed a confidence prediction model for Ad campaigns which helps to spend high value on those has good score.

Platforms and frameworks I have worked with :

- *Deployment tools & services* : GCP , Docker, Airflow , Mlflow, Vertexai , cloud composure
- *Frameworks/Libraries* : Sklearn , Pandas ,numpy , bigquery
- *Languages* : Python, CCP , Sql , Pyspark , bashscript , ymal

Impact:

Helps in Segment targeting and trigger targeting for Showing ads to groups of users with consideration certain set of attributes.

Ad Bid Optimization

Realtime bidding model to predict the price for the ads that's published in social medias .

- ✓ Under development using reinforcement learning Markov Decision Process (MDP) and model shall generates data per hour of the impressions performed during the real-time bidding in PoC .
- ✓ This model is powed by ADS Data hub which is large data Platform from Google ads and Its aim to reduce 30% money in spending ads by decreasing the noise on non valuable ads in real time.

Tamilnadu e-Governance Agency, Chennai - AI/ML ENGINEER **(March 2021 – May 2022)**

Tamilnadu State Government IT Department -Centre of excellence in emerging technologies(CEET)

Roles & Responsibility

- ❖ Ideating department/government use cases and convert them into a well-defined key problem statement to **End to end scalable AI solution** followed by **Prototyping, Testing and Productization. – including analysis, planning, model, design, develop, code, test, debug, documentation .**
- ❖ Leading and handling the **critical deliverables, benchmarking solutions, driving Key Metrics** and ensuring projects are usable to state peoples.
- ❖ Collecting datasets required to **develop predictive models** for solving government problem statements.
- ❖ Developing Computer vision models , Natural Language Neural networks/Models and predictive algorithms to derive insights and business value from data.
- ❖ Making model as api service using Django , Flask , Fastapi as deployable solution in web application and mobile solution.
- ❖ Maintaining TESLA GPU Servers and whole organization project code in GITLAB . Automating the backup service with cronjob in servers.

Projects :

Epaarvai – AI Cataract deduction application

An AI intelligent Android mobile application to provide immediate screening for cataract on an individual's eyes and if found positive for cataract then share the details with the district and the state authorities.

- ✓ Architected and Developed the entire AI Stack , Backend with **Lightweight Prediction** services using **fastapi framework** and provides production-ready code and contributed in integration with flutter app
- ✓ Created Entire training *dataset* was from *scratch* and developed standard operating procedures (SoPs) for collecting the **dataset** from *over **700+ patient***. ***Created a labelling platorm and annotated*** the datasets.
- ✓ Combined a state of art AI Models like on **CSPDarknet53, SPP, PANnet, and YOLOv3** .The model is Prediction **is 98.5% Accuracy** sensitive.
- ✓ Built and pipelined a robust scalable ai architecture and Deployed in the ***AWS server and Dockized*** the inference and Django api .

Platforms and frameworks I have worked with :

- *Deployment tools & services* : AWS , Docker, DVC , Mlflow, Kubeflow.
- *Frameworks* : Pytorch, Django, Ngnix ,Fastapi , gunicorn
- *Languages* : Python, HTML , Flutter , React JS
- *Build* : Cmake,Cuda
- *Database* : PostgreSQL.

Impact:

The E-Paarvai app is currently being used by the Tamil Nadu State Control Blindness Society (TNSBCS) and the solution has been rolled out to more than 32+ districts and have screened more than 20000+ patients in the various rural areas of Tamil Nadu.

Awards & Recognitions :

- ✓ The innovation had a huge social impact won the Nasscom ***AI Gamechangers Award in 2021*** as one of the best innovations in AI!
- ✓ 75@75 India's AI Journey report published by Digital India in that our project secured place in ***top 3*** .

About project :

Youtube link : *ePaarvai - AI Cataract Application*

Uzhavan Predictor – Pest and Disease detection & Recommendation System

An AI intelligent solution to provide crop diagnosis and recommendation for fertilizers by identifying the pest & disease which has affected crop and sending the remedial measures using the photo.

- ✓ Architected and Developed the ***End to End entire ML and CV Solution , Created a Flask api*** point for Web application stack and ***rendered the webpage*** to display the results.
- ✓ Build a model, which can ***classify between healthy and diseased*** crop leaves and also if the crop have any disease, predict which disease is it.
- ✓ Used a ***ResNet – 50 Model*** , results ***AI Prediction of 95.5%*** Accuracy sensitive measured using Classification Reports, ***Confusion Matrix, AUC Score.*** More than 20+ variety of crop classes are integrated.
- ✓ Developed helper functions like ***data loaders*** been used and ***Data augmentation*** are followed for better outcomes.
- ✓ Build various ML Algorithms – ***Decision Tree , Guassian Naïve Bayes , SVM , Logistic Regression ,Random Forest , XGBoos***t and compared the accuracy using ***TF Board*** .Best model is Served as the ***pickle file*** .
- ✓ Performed a various ***pre-processing*** techniques for cleaning ***data cleansing***.

Platforms and frameworks I have worked with :

- *Deployment tools & services* : Flask , TF executor, TF jobs, TF Serving , DVC .
- Libraries Used : Scikit-learn , Pandas , Numpy ,Seaborn, matplotlib
- *Frameworks* : Pytorch, Fastai , uvicorn , nginx
- *Database* : Mysql.
- *Languages* : Python, HTML

Impact:

With No of cost farmers can find the

Awards & Recognitions:

75@75 India's AI Journey report published by Digital India in that our project secured place in top 15.

About project:

Youtube link : Uzhavan Recomendation system | Uzhavan Pest Deduction

Facial Recognition System (Android & Edge Device)– For Schools , Rehabilitation (Refugees) & Medical institutes

Developed a scalable and easy deployable android monitoring system for the migrants for identity tracking with their Geolocation.

- ✓ Architected Centralized face attendance system, pipelined the stack , introduced ERP features in the Web dashboard .
- ✓ Developed the ***Android Based Inference for Face recognition using TFlite , Opencv for Java , CameraX.***
- ✓ Used a ***Light weighted Mobile optimized models -*** Facenet , Arcface/Deepface , RetinaNet and results 99.5 Percent of True positive for Localizing the Indian faces.
- ✓ Optimized **JWT tokens** to sync from client device to central servers in backend.
- ✓ Novelty of the application is not save the users picture and by using only the embeddings of users to deduct them.
- ✓ Training of user data will occur locally in the android device without any GPU dependencies separately for each users takes only 30s for creating the model / embeddings.
- ✓ Deploys solution in various premise depends on requirements from scratch .
- ✓ Manually integrate PI 4 – Equipped with the camera and necessity components to the servers and train the users .

Platforms and frameworks I have worked with :

- *Deployment tools & services* : Docker , Bash scripting , DVC , PI 4 , Jetson Nano , Ubuntu servers, XAMP.
- Libraries Used : Opencv,Dlib , Sckitlearn ,Mlflow

- *Frameworks* : TFlite , Opencv Java , Django, Android Firebase
- Inference : Android Camera X, PI Inferance
- *Languages* : Python, HTML , Angular JS , PHP, Bash script
- *Database* : MySQL, SQLlite

Impact:

The technology is currently being used by the Tamil Nadu Government Premise for monitoring the contactless attendance and for schools tracking the illiteracy count of the students over period.

Awards & Recognitions

75@75 India's AI Journey report published by Digital India in that our project secured place in top 15.

About project :

Youtube link : FRAS -AI Solution

ASHOKLEYLAND - INDIAN DEFENCE SYSTEM , *Hosur - Jr. NPD ENGINEER* (May 2018 – Oct 2020)

- ✓ Developed a IOT based object deduction for Parts Inwarding store using **SSD and Kalman filter** to extract the part number.
- ✓ Used **google coral** device to process high speed image recognition as input external device
- ✓ Create and maintain optimal data pipeline architecture, assemble large data that meet functional / non-functional business requirements.
- ✓ Identify, design, and implement internal process improvements: automating manual processes, optimizing data delivery, re-designing infrastructure for greater scalability.

Libraries Worked with: Pandas, TensorFlow , TF IDF,Opencv , Scikit-learn, TQDM

Skillset: K-NN, Logistic Regression Classifier, Clustering, Transfer Learning, Resnet , VGG 18

Responsibilities

- *Data labelling*
- *Backend API Creation*
- *Data Augmentation*
- *Ensuring SQL Server*
- *Deployment of models*
- *Monitoring model*
- *FMEA discussion*
- *CFT Member*
- *Achieving KRA / KPI monitoring*

Recognitions

- ✓ Successfully Completed **VQAW Customer Audit** in Supplier end for **Load body & Driver Cab** - *M/s Prabha 1 & 2*
- ✓ Reduction in inspection lead time through introduction of **checking templates, gauges** for complex specifications.
- ✓ Successfully Implemented more than **100 Nos of CAPA** for identified problem and various defects.
- ✓ Introduced **part identification system** for machining & sheet metal parts for better tracking in the field.
- ✓ **Trained quality inspectors** and **Shop floor operators** on the quality requirements and method of inspection.

Research Experience

24th National Conference on e Governance, **Central Gov India – Author/ Researcher** (Jan 2022)

Awarding winning research paper from TN State government in theme of ***" India's Techade: D&gital Governance in a Post Pandernic World"***

- ✓ The architecture is based on the object detection algorithm that divides images into a grid system.
- ✓ A deeper design gives tenfold improved expressive potential as compared to standard shallow models
- ✓ The network has 16 convolutional (conv) layers, 2 up sampling layers, 4 concat layers, and a soft max classification layer, and can be optimised on an objective function.

University of Brunei Darussalam, **Brunei,** NLP **Research assistant** (Dec 2020 –Feb 2021)

- ✓ Developed a time series model as a proof of concept to recognize patterns using RNN for stock market. Proposed a methodology to solve the inadequate data problem of knowledgebase for Goal Oriented Accuracy, using transfer learning, attention mechanism and improved the performance by 10-15%.

Work and Extra Curricular Experiences in the AI

HIKE INSTITUTE *-Entrepreneur – Educator/Mentor* (Dec 2019 – Present)

Teaches the students /aspirants for Building their career Field of AI/ML. Helped 40+ peoples for transiting into this tech.

iNeuron Intelligence Pvt. Ltd , **Bangalore , COMPUTER VISION ENGINEER** (Oct 2020 –Feb 2021)

Ultra-Fast Structure-aware Deep Lane Detection –

- ✓ Successfully built live lane detection using deep learning neural networks, computer vision techniques, and OpenCV.
- ✓ Our model using Resnet-34 backbone could achieve state-of-the-art accuracy and speed. A light weight Resnet-18 version of our method could even achieve 60 fps.
- ✓ Inherent structure of lanes, a classification-based network prediction and auxiliary segmentation task are illustrated in the green and orange boxes. The group classification is conducted on each row anchor.

Libraries: NumPy, Pandas, Pytorch, Scikit-learn, OpenCV , TQDM , Python
Skillset: Neural Networks,CNN, Clustering, Transfer Learning, Resnet , VGG 18

Sentiment Analysis with Deep Learning using BERT **(Aug 2020 –Sept 2020)**

- ✓ Clean and pre-process text dataset.Split dataset into training and validation sets using stratified approach.Tokenize (encode) dataset using BERT toknizer.Design BERT finetuning architecture, Setting Up Optimizer and Scheduler.Evaluate performance using F1 scores and accuracy.Finetune BERT using training loop.

EDUCATION

DEGREE	NAME OF INSTITUION	YEAR OF PASSING	PERCENTAGE
B.E Automobile Engineering	Hindusthan college of engineering & Technology, Coimbatore	2018	7.78(CGPA)
HSC	Cheran Higher secondary school, Karur	2014	82.25%
SSLC	Cheran Higher secondary school, Karur	2012	80.25%

(VARUNSAAGAR.S)