NIRMITK TRIPATHII
AI/Machine Learning Engineer

Mobile: 9528006968
Email ID: nirmitktripathii@yahoo.com
Address: Gurgaon, IN
DOB: 10/03/1995
LinkedIn: _https://www.linkedin.com/in/nirmitk-tripathii-553a92158/_
Github: _https://nirmitktripathii.github.io/PortfolioSite/_

With a deep understanding of Data Science, Machine Learning, and Data Analysis gained through a Masters in Data Modeling and Simulation, my goal is to secure a challenging Machine Learning Engineer position in the Computer Software/Engineering industry. I aspire to employ my technical skills, project experience, and passion for exploring data patterns, to not just revolutionize the way businesses operate, but also satisfy my intellectual curiosity. I aim to contribute to a progressive company aligned with my growth trajectory and grasp opportunities that let me leverage my skills while pushing me to learn further.

➤ PROFESSIONAL SUMMARY

- Certified Data Scientist with over 2 years of industry experience, specializing in AI and ML techniques.
- Substantial experience using Python, Selenium, Scrapy, Scikit-learn, Pytorch, Tensorflow and Gen-AI models, especially Google-Gemini and more to manipulate data and draw insights from large datasets.
- Expertise in applying data-driven, machine learning approaches to business challenges, having successfully implemented models in commercial projects.
- Proven proficiency in designing optimized data scraping processes and building business solutions based on Neural Networks (ANN), Natural Language Processing (NLP) and Large Language Models (LLMs)
- In-depth understanding and application of data analytics on a variety of fields ranging from football for performance predictions and comprehensive sport analysis, to speech-to-text conversion for customer-sales call summarization and extracting insights from the conversation to agriculture for disease classification of crops based on images of leaves
- Accomplished in successfully delivering Data Science and AI/ML based projects, meeting quality parameters and time-lines.
- Expert practical knowledge in business application of open source LLMs like BERT, Open-Llama, Llama-2 & 3, phi-2 & 3, Gemma-2 etc.

➤ WORK EXPERIENCE

AI/ML Engineer
Blu Parrot   (Jan 2024 - Present)

Lead discovery processes with project stakeholders to identify the business requirements and the expected outcome

Contribute with other members of the team for the design-related and algorithmic innovations and abstracts

Contribute in Building models for creating, maintaining and using knowledge base consisting of semantic relationships

Be able to work on databases, scientific computing, predictive modeling, and data analysis for building machine learning components and implementing algorithms on the back end

Focus on data analytics and produce results and recommendations using supervised and unsupervised Machine Learning and/or AI (LLM and Transformer based architecture)

Make Story board or presentation on the topics of interest to present to customer

- Created a **portal for transcribing (automatic speech recognition) of recorded audio calls in Hindi language using STT Hindi Conformer-CTC model**, and performing **natural language processing and understanding on the transcribed text** to gain **valuable insights** from the conversation and **recommended future plan of action**, and **extracted call summary** using Google Gemini.
- Performed **finetuning of multilingual XLM-RoBERTa and IndicBERT on sentiment analysis** of social media chatter related to Indian General Elections 2024.
- Performed **data scraping** and **extracted product information** on **android grocery apps** and **perform data mapping** and **data analysis** on the extracted product information.
- Created a **pipeline to read, clean and convert scanned pages from a PDF into text searchable format using Microsoft Azure OCR and OpenCV.**
- Created a **portal to translate any text in 22 officially designated Indic languages (including English)** from and to, any other remaining 21 Indic languages (including English) using **IndicBERT** and **Indic-TransV2 model.**
- **Extracted** various categories and types of **data** related to **real estate, construction, property development and investment.**
- Currently working on building a **financial regression model** to predict productivity, inflation and employment metrics based on historical data and various other factors related to UK real estate, property and construction industry.

SOFT SKILLS

Problem-solving
Time Management
Analytical Thinking
Team Collaboration
Communication

TECHNICAL SKILLS

Skill	Level
Tableau	●●○○○
Python	●●●●●
Data Mining	●●●●●
Large Language Modelling	●●●●●
Selenium	●●●●●
Scrapy	●●●○○
Deep Learning	●●●●●
Machine Learning	●●●●●
Natural Language Processing	●●●●●

CORE COMPETENCIES

Neural Networks and Deep Learning
Data Science
Back-End Development
Large Language Modelling
Web Scraping
Sports Analytics
Machine Learning
Data Analysis
Natural Language Processing

EDUCATION

M.Tech: Modeling And Simulation
Savitribai Phule Pune University
Sep 2019 - Feb 2022

B.Tech: Electrical Engineering
Pandit Deendayal Petroleum University
Jul 2014 - Jun 2018

LANGUAGES

English, Hindi

HOBBIES

Engaging in European History and Ancient and Classical Indian History,

Studying Astrophysics and Quantum Mechanics,

Reading - Particularly, the works by Stephen Hawking, Richard Dawkins, Erwin Schrodinger, Dan Brown and Jeffrey Archer,

Exploring new and fascinating aspects of mathematics everyday

Achievements:

Got promoted as Project Lead to a Fintech Portal development project related to real estate in UK

Subject Matter Expert: AI/ML
House of Coutons   (Nov 2023 – Jan 2024)

Conducting research in computer science and AI/ML domain and sending proposals for researchers and clients.
Delivering research papers, and dissertations along with working models to client on topics related to AI, Machine Learning, Deep Learning and Natural Language Processing.

- Delivered 3 research papers and 1 dissertation on different research topics related to AI/ML field.
- Completed and delivered ML model for 1 research paper and 1 dissertation with complete literature review, EDA, model building and inferences.

AI/ML Programmer
HRBOTICS   (Apr 2022 - Mar 2023)

I managed the company's AI/ML activities focused mainly on the acquisition, storage, and analysis of data. The work leaned heavily on writing automations to facilitate HR practices.

- Acquiring data of all the registered commercial outlets in NCR from 2010 to 2021 using data scraping with multi-threading.
- Automating the entire workflow of job posting and sharing as well as applicant hiring in Linkedin using headless Selenium an BeautifulSoup, creating an ATS protoype.
- Design reusable libraries and application programming interfaces (APIs) for streamlining and automating ATS.

Achievements:

Created an automated ATS system for HR recruitment using Selenium and BeautifulSoup.

Football Data Analyst
Boomer 11   (Nov 2021 - Mar 2022)

My primary role at Boomer 11 involved the acquisition, analysis, and prediction of football match statistics to provide insightful player and team assessments and to predict potential outcomes.

- Prediction and analysis of key battles (a defender vs a striker, midfielder vs midfielder and winger vs full back) between the players from both the sides.
- Regular acquisition of real time statistics from matches. Application of the acquired data into performance analysis of individual players and teams.
- Video podcasts discussing the latest happenings in the footballing world and analysing important matches, transfer strategies, and playing tactics.
- Analysis and prediction of important players to look out for in the next match based on the percentile and raw stats acquired for the players from previous matches.

Achievements:

- Created models to quantify ratings of football players based on estimated probabilities of their in game actions.
- Created an Expected Goals model that gives probability of a shot taken in a football match being a goal or not.

⟩ PROJECTS

AI Call Notes
- Developed a robust portal leveraging the STT Hindi Conformer-CTC model for transcribing recorded Hindi and English audio calls, achieving State of the Art (SoTA) results for Hindi audio transcription.
- Integrating advanced natural language processing (NLP) to analyze the transcriptions and extract actionable insights from conversations.
- Additionally, utilized Google Gemini to generate comprehensive call summaries and provide strategic recommendations for future actions.
- Achieved zero-shot 92%+ accuracy

AI Indic Translator
- Developed an advanced translation portal utilizing the IndicBERT and Indic-TransV2 models, enabling seamless translation across 22 officially designated Indic languages (including English). This system facilitates translation from any language to the remaining 21, offering robust multilingual support for diverse linguistic needs.
- Can successfully translate to and from the following Indic languages:
- Assamese, Bengali, Bodo, Dogri, English, Gujarati, Hindi, Kannada, Kashmiri, Konkani, Maithili, Malayalam, Manipuri (Meitei), Marathi, Nepali, Odia, Punjabi, Santali, Sindhi, Tamil, Telugu, and Urdu
- Achieved zero-shot 94% accuracy

Twitter Sentiment Analysis Tool
- Built a Twitter sentiment analysis tool that classifies a tweet into positive or negative sentiment.
- Achieved an overall accuracy of 87% on unknown test data using TF-IDF token vectorizer.
- Developed a model app on Flask with a frontend which takes a tweet as input, sends it to the Twitter Sentiment model in the backend and displays the result back in the frontend as output.

Image Classifier Using ImageNet CNN Model
- Created an Image classification Deep Learning Model using an ImageNet CNN network architecture (VGG16) trained on pictures of different unhealthy plants, leaves and fruits and classify them according to their different diseases.
- Achieved an accuracy of 84% on unknown test data.
- Created a web application using Flask which takes input images and classify the species, health status and disease of the plant, if it classified as not healthy.

Modeling Shot Quality Using Expected Goals Metric
- Developed an ultra fast data scraping code to web scrape shots data from Understat.com website that reduced the time for webscraping shots data from the site from 8000+ seconds to 250 seconds.

- Developed a Random Forest (RF) and Logistic Regression (LR) model to predict and classify a shot taken by a player as a goal or no goal depending on the distance and angle made by the shot location from the goal mouth as well as whether the shot was taken by strong foot/weak foot/header.
- Got 89.77 % accuracy with RF Classifier model without hyperparameter optimization and 90.55% accuracy with hyperparameter optimization. xGBoost Classifier achieved 90.55% accuracy as well but had a poor f1- score in classification of goals as compared to RF classifier.

EXPLORATORY DATA ANALYSIS, & DEVELOPING A VAEP MODEL FOR PLAYERS & PREDICTIVE MODEL FOR TOP 5 EUROPEAN LEAGUES IN FOOTBALL
- EDA aimed at studying the football dataset, to analyse, extract information from it and make important conclusions based on the data, primarily on the different styles of plays in different teams, finding weaknesses and strengths of the teams and assess the ways of measurement and improvement of the team performance.
- For predictions, used multiple output regressor with different regressors such as Random forest regressor (RF), Support Vector Regressor (SVR), to predict the points table. Got 67% accuracy in Top 6 prediction and 50% accuracy in Bottom 4 prediction for 2019-20 season.
- Created a GUI wherein we display the goal plots in which all the previous actions leading to a goal are displayed and top 10 players, of each league, according to VAEP (Valuing Actions based on Estimated Probability) rankings are also displayed.

➤ CERTIFICATIONS

Masters Program – Data Scientist (2023)
Simplilearn
Mad About Sports (2021)
Football Data Analytics

➤ ACHIEVEMENTS

- Developed an Expected Goals model, using RF Classifier, which had 89.77% accuracy predicting shot outcomes in football matches.

  Publication: http://dx.doi.org/10.13140/RG.2.2.35007.73128

- Developed a robust portal leveraging the STT Hindi Conformer-CTC model for transcribing recorded Hindi and English audio calls, achieving State of the Art (SoTA) results for Hindi audio transcription, and integrating advanced natural language processing (NLP) to analyze the transcriptions and extract actionable insights from conversations. Additionally, utilized Google Gemini to generate comprehensive call summaries and provide strategic recommendations for future actions.

- Created a portal to translate any text in 22 officially designated Indic languages (including English) from and to, any other remaining 21 Indic languages (including English) using IndicBERT and Indic-TransV2 model.