ASHUTOSH MISHRA

Github, linkedin, Kaggel
CONTACT +919968514930 **amishra527@gmail.com**

EDUCATION

CDAC , Noida — Diploma in Artificial Intelligence.
March 2021 - Sep. 2021

- Specialized in Machine Learning, Deep Learning, NLP, CV.
- Build and implemented Lots of ML, DL algorithm (Regression, SVM, CNN, GAN, YOLO, ANN, Random Forest, Classification, K-Means, kNN, GBM, Naive Bayes, etc)

Himalyan Garhwal University, Pauri — MSc.
July 2017 - July 2019

- Specialized in computer science and scored 70% Marks.
- Created a Parkinson's disease detection algorithm with a Primary goal to achieve accuracy of 94%.

Imr, Ghaziabad — BCA
July 2008 - July 2011

- Specialized in Computer Application and Scored 62% marks.
- Build a project in the school management system in VB.

EXPERIENCE

DEV INFORMATION TECHNOLOGY LTD,
Deployed on the **Government Project** in **NIC Headquarters, CGO Complex New Delhi**,
Currently working as **Sr. Software Developer** in the **Artificial Intelligence Research Department(AIRD)**
Projects:-

- **Nvidia NEMO Translation(Nvidia AI Platform)** For **Panini translation Service of Indian Government.**
  - **Goal:-** To correctly translate English to Hindi (Indic languages Assamese, Bengali, Gujarati, Kannada, Malayalam, Marathi, Oriya, Punjabi, Tamil, Telugu)
  - **Technology:- Python, Transformer sequence-to-sequence architecture, pytorch, Nemo, FastApi, Kubernetes, docker, Locust, Nvidia RIVA API Server.**
  - **Operations:- Collect the data collection, perform cleaning, Normalization, Tokenization, Training and applying BPE tokenization, Batching, Bucketing and padding.**
  - Tarred dataset for large corpora, trained model from scratch, test the performance of the api by locust.
  - **Translation Accuracy:-** 58%.

- **Finetune IndicTrans Translation Service From AI4Bharat for Panini translation Service of Indian Government.**
  - **Goal:-** Train the model in accordance with Government data for correct output in **11 indic languages.**
  - **Technology:- Python, Pytorch, transformer 4x, Kubernetes, docker, FastApi, locust.**
  - **Operations:-** Build API to consume the service and test it for the number of users it can handle.
  - **Result:-** Got the good results for the general translation and have a good response from the **Departments/Ministry**.
  - **Translation Accuracy:-** 78%.

- **Finetune IndicTrans 2 Translation Service From AI4Bharat for Panini translation Service of Indian Government.**
  - **Goal:-** Train the model in accordance with Government data for correct output in **22 indic languages.**
  - **Technology:- Python, Pytorch, transformer 4x, Kubernetes, docker, FastApi, locust.**
  - **Operations:-** Build API to consume the service and test it for the number of users it can handle.

- **Result:-** Got the good results for the general translation and have a good response from the **Departments/Ministry.**
- **Translation Accuracy:-** 88%.

- **Panini translation Service** built a chrome extension which can translate the whole webpage into any target language.
  - **Goal:-** To utilize the API of panini service and translate the whole web-page.
  - **Technology:-** **javascript, fastAPI.**
  - **Operations:-** Build and test the chrome extension on multiple websites. Extracted the text from the web-page, stored, sent to api, received the result and replaced with original text.
  - **Result:-** Performance was satisfactory(tested for both linearly and in batches) as not able to resolve the UL tags.

- Built a **HTML/XML to HTML/XML API service** which can translate the whole web-page into any target language.
  - **Goal:-** Send the HTML page, select the translated language and receive the translation page.
  - **Technology:-** **Python, Javascript, HTML, fastAPI and Beautifulsoup.**
  - **Operations:-** Extracted the text from the HTML/XML page, stored, sent to api, received the result and replaced with original text.
  - **Result:- Achieved the goal. Tested for Both one by one request or in batches.**

- Built a **PDF to Doc's Converter English to Different Indic Languages using Ai-Panini Translation service** which can translate the whole text content of a page into any target language.
  - **Goal:- Read and Extract the PDF's content and save a copy in a Doc File and** Send the Text to the translated Service and receive the translation Output.
  - **Technology:-** **Python, Tesseract, fastAPI and Opencv.**
  - **Operations:-** Extracted the text from the PDF, stored, sent to api, received the result and stored in the Docx file along with the original Extract text from the pdf.
  - **Result:- Achieved the goal. Tested for Both one by one request or in batches. With some formatting issues.**

- Built a **PDF to PDF's Converter from English to Different Indic Languages using Ai-Panini Translation service** which can translate the whole text content of a page into any target language.
  - **Goal:- Convert and Extract(OCR) the PDF's content and save it into an image(PNG,JPG)(Converted) File and** Send the Text(Extracted By OCR) to the translated Service and receive the translation Output.
  - **Technology:-** **Python, Tesseract, fastAPI and Opencv.**
  - **Operations:-** Extracted the text from the PDF, stored, sent to api, received the result and stored in the Image file later converted into PDF along with the original Extract text from the pdf.
  - **Result:- Achieved the goal. Tested for Both one by one request or in batches. With some formatting issues.**

Genshin Impact Anime Faces.
Goal:- create images that look like photographs of human faces
Technology:- Python,DL Algorithm GAN

Financial Sentiment Analysis.
Goal:- To correctly classify the statements(Financial)

Technology:- Python, NLP(Bag of Words), ML Algorithms(Naive Bayes,MultinomialNB, XGBoost, CatBoost)
Accuracy:- 68%

India's mineral ores.
Goal:- Use Geoplot to correctly identify the different types of ore concentration.
Technology:- Python,geopandas, geoplot, EDA

UNIVERSITY PROJECTS

DNA Sequence classifier using CNN —
Goal:- To correctly classify the DNA Sequence()
Datasets:- two datasets Splice(Extron-Intron, Intron-Extron, None), H3(wrap around Histone, None)
Technology:- Python, NLP, DL Algorithms(CNN)
Accuracy:- train(99.3%, 85.9%), test(97.5%, 82.4%)

Parkinson's Disease Detection —
Goal:- To correctly classify the Patients.
Technology:- Python, DL Algorithms(Multi-Layer Perceptron, Tensorflow)
Accuracy:- 91.3%

CERTIFICATION

Python for Machine Learning. Basics of Exploratory Data Analysis. Probability for Data Science. Statistical Method for Decision Making. Visualization Python, Tableau. Predictive Modeling and Analytics. Spark:Pyspark.

ADDITIONAL

Technical Skills: Python, Data Analyst, Machine Learning, Deep Learning, NLP,OpenCV, Spark,PowerBI, Tableau, Excel, SQL.
Languages: HINDI, ENGLISH.