Chandan Purbey

Data Scientist

+91 9126452390 chandanpurbey83@gmail.com Kolkata linkedin.com/in/chandan-purbey/

SUMMARY

- Bachelor's Degree in Computer Science & Engineering.
- **4+ years** of hands-on experience in **Data Science, Machine Learning, Deep Learning, Convolutional Neural Networks (CNN)** and **Natural Language Processing (NLP)**.
- Proficient in **Python**, with strong expertise in **Flask** and essential libraries including **Numpy, Pandas, Matplotlib, Seaborn, Scikit-learn, TensorFlow, Keras, NLTK, Spacy, Transformers, BERT**, etc.
- Skilled in data storage and SQL queries, utilizing **PostgreSQL.**
- Hands-on experience on cutting-edge tools and technologies such as **OpenAI, Langchain, etc.**

TECHNICAL SKILLS

Languages: Python
Libraries: Pandas, Numpy, Scikit-Learn, OpenCV, Matplotlib, Seaborn, Spacy, Keras, Tensorflow, OCR, NLTK
Machine/Deep Learning: Xgboost, CNN, Clustering, Support Vector Machines(SVM), KNN, Decision Trees, Random Forest
Cloud Computing: S3, AWS Glue, Azure Pipeline
Database: PostgreSQL, MySQL
Statistics: Ensemble Techniques, Hypothesis Testing, Standardization, Normalization
Frameworks: Flask, Django
Tools: Jira, Git, Github, Github Actions, Docker

PROFESSIONAL EXPERIENCE

Consultant (Data Scientist) — Jul '24 - Present
Capgemini — Kolkata, IN

Project: GenAI Chatbot

- Developed and maintained a data processing platform with ingestion from multiple sources, including AWS **S3**, and implemented efficient data retrieval from ElasticSearch.
- Automated workflows using Azure Pipelines and AWS Glue jobs.
- Monitored system performance through CloudWatch and managed code repositories using GitHub. Integrated OpenAI solutions and AWS Lambda for API-driven tasks and automation.

Key technologies and tools: **Azure Pipelines, AWS Glue, CloudWatch, GitHub, OpenAI**, AWS Lambda, API integration, **ElasticSearch**.

System Engineer (Data Scientist) — Oct '20 - Jun '24
Tata Consultancy Services (TCS) — Guwahati, IN

Project: Document Review Automation

- **Enhanced Blank Field Detection with NER models(NLP) & CNN:**
  - Detection of blank fields in documents using a combination of **NER**, **CNN**, and **OCR** tools such as EasyOCR, Keras OCR, and Tesseract OCR for both scanned and typed documents.
  - Implemented NER models using **Spacy** and custom entities on a corpus of around 10,000 annotated sentences from Clinical Trial documents, tailored to specific business requirements, achieving an accuracy of 91% and comparing it against benchmarks.

- **Image Classification & Blur Detection:** Readability
  - Designed and trained image classification models using **Convolutional Neural Networks (CNNs)**, including **pre-trained models** like **VGG16, VGG19, ResNet**, and **Xception**, for feature extraction and categorization.
  - Executed **transfer learning** and fine-tuning to achieve an accuracy of 96.66%, utilizing frameworks like **Keras**, **TensorFlow**, and **OpenCV**.

.

Project: Sentiment Analysis of Drug Adverse Events: Insights from Medical Comments

- Conducted sentiment analysis on drug adverse event reports to evaluate the polarity of effects, utilizing comments from medical professionals and patients.
- Employed **NLP** feature engineering techniques including **bag of words**, n-grams, **TF-IDF**, etc and **transformer**-based **LLM** model like **BERT** to capture crucial sentiment-related information from the data.
- Achieved an impressive F1-score of 94% through rigorous evaluation and fine-tuning of sentiment analysis models.

Project: Labeling/Classification of Drug Administration Routes

- Designed and deployed a robust **multi-class** classification model to accurately categorize drug administration routes, enhancing efficiency and accuracy in pharmaceutical data management.
- Designed and Implemented **XGBOOST** model in **multi-Label classification** model and classified data in multiple categories with **Model F1-Score: 91%**

Skills

- Machine Learning Methodologies • Optimization Techniques • Text Classification • Data Mining & Analytics • Predictive Models
- Predictive & Statistical Modelling • Sentiment Analysis • Tensorflow • Data Analysis • Data Visualization • Statistical Algorithms
- Predictive Modelling & Analytics • Programming • Natural Language Processing • Statistics Methods • Statistical Modeling
- Team Coordination & Leadership • Deep Learning • Semantic Analysis • Data Manipulation • Machine Learning Frameworks

EDUCATION

B.E in Computer Science Engineering Aug '16 - Jul '20

Gauhati University Guwahati, IN

- **Assam Engineering College**

Higher Secondary (Science) Aug '13 - May '15

Assam Higher Secondary Education Council Guwahati, IN

- **B. Borooah College**

CERTIFICATIONS

- Data Science Professional Certificate
- Cutshort Certified Python – Advanced
- Python
- SQL
- Problem Solving
- AWS Developer – Associate