Shashank Bhatia

Data Scientist

shashankbhatia15@gmail.com | +91 8368816137 | github.com/shashankbhatia15 | linkedin.com/in/sbhatia15

Profile

Driven and accomplished data scientist specializing in Machine Learning, Deep Learning, and Natural Language Processing (NLP). Demonstrated proficiency in Python and SQL, with expertise in utilizing advanced frameworks including Keras, TensorFlow, Scikit-learn, NLTK, OpenCV, as well as platforms like AWS (EC2, Sagemaker), Jupyter, and Azure. Recognized for pioneering and executing cutting-edge NLP models to extract valuable insights from textual data.

Skills

Languages

Python, SQL

Platforms and Frameworks

AWS EC2, Sagemaker, Streamlit, Jupyter, Azure, Keras, Tensorflow, Scikitlearn, Selenium,SciPy, NLTK,OpenCV

Data science

Linear & Logistic Regression, Decision Tree, Random Forest, XGBoost, KNN, Kmeans, CNN, RNN, LSTM, ML, DL

Others

Lang chain, LLM, Prompt Tuning, Excel, Professional communication, Storytelling, Problem solving, Critical Thinking, Strategy, Prompt Engineering

Professional Experience

The Smart Cube

Aug 2022 – present

Senior analyst – Data science

PDF QnA using LLM

- Spearheaded the development of a robust Document QnA pipeline using **LangChain** to extract relevant information from 50+ PDFs with its text, summary, meta data and tables as input.
- Implemented open-source Embeddings, tokenizers and **large language models** to enhance natural language processing capabilities.

Text-to-SQL based QnA model

- Innovated a tool to facilitate data-driven decision making, empowering users to extract valuable market information from a comprehensive tabular dataset of over 20,000 commodities.
- Fine tuned a **BERT** based classification model that discerns user intent from the extracted query components, predicting the most appropriate SQL query with a **0.93 F1 score**.
- Engineered a parsing mechanism to extract essential components from user queries, including commodity name, region, and other relevant parameters using NER.
- Instituted a custom **NLQ** engine to translate user input into executable code for 6 databases, enabling **SQL** query generation.
- Leveraged **prompt engineering** using **Langchain** to generate tailored summaries from **ChatGPT** meeting the specific needs of users.

News Risk Classification

- Developed two models to predict risk ratings and news category, performing multi-class and multi-label classification.
- Applied **NLP** techniques such as data cleaning, Tokenization and lemmatization for text column processing.
- Utilized the Synthetic Minority Over-sampling Technique **(SMOTE)** to address imbalanced data (98% and 2%).
- Trained models using **logistic regression** and **XGBoost classifier,** achieving an **recall** of **0.82** and **0.84** respectively.

Wipro Technologies

Sep 2017 – Jul 2022

Project Engineer

Clause Extraction and Classification Model

- Formulated a solution to extract cargoes that are allowed/not allowed from 20000+ PDF files.
- Eliminated strike-through text from the text and extracted 2 exclusion clauses from raw text using **Regex**.
- Built an **XGBoost** model that identifies correct exclusion clauses with a **0.97 accuracy**.
- Fine tuned a custom-named entity recognition **(NER)** model using **Spacy** to extract 36000+ commodities.
- Built a **Roberta** classification model using MLM tokenization, classifying the cargoes from a text with a **precision** of **0.89**.

Design Data Analysis and Failure Prediction

- Executed exploratory data analysis **(EDA)** on a large dataset of automobile design data, uncovering **crucial insights** and patterns that contributed to identification of design failure factors.
- Developed a **Logistic Regression** that accurately predicted design failures, achieving a **precision** of **0.81**.

Vehicle design and documentation

- Modification in **Design** and **Digital Product Modelling** on **Catia**.
- Orchestrated **Python scripts** to enhance and streamline Quality Control (QC) processes, resulting in a **40%** reduction in time required.

Education

B.Tech Mechanical and Automation Engineering

2017

Amity University, Noida

CGPA - 7.89

Certifications

Azure AI Fundamentals (Microsoft) | **Deep Learning A-Z** (Udemy)

Awards and Achievements

High Honors (The Smart Cube , 2023) | **Kudos Award** (The Smart Cube , 2022)

Victory League [Extraordinary Commitment] (Wipro Technologies - 2019,2021)