

ARIZ KHAN

✉ aarizkhan2580@gmail.com ☎ 9654396200 🔗 linkedin.com/in/arizkhan 🌐 portfolio/arizkhan

Professional Experience

Software Engineer, Appolo Computers Pvt Ltd

May 2023 – Present | New Delhi

- Hands-on experience in Natural Language Processing, specifically leveraging transformer models like BERT, BART, Mistral and GPT-3.5 etc.
- Utilizing Elasticsearch, enabling efficient and precise search capabilities within PDF documents.
- Contributed to building systems that streamline information retrieval from extensive document repositories.

Data Science Intern, AI Variant

Aug 2022 – Apr 2023 | Hyderabad

- Hands-on experience in developing a job recommendation system by implementing web scraping with Python (BeautifulSoup, Selenium) and applying NLP techniques (spaCy) to analyze LinkedIn data, optimizing personalized job matching.
- Proficient in utilizing NLP models like spaCy and NLTK for sentiment analysis and text extraction, enhancing guest experience through insights from hotel reviews.
- Expertise in building predictive machine learning models using Random Forest, Gradient Boosting, and XGBoost to identify customer churn, leveraging advanced feature engineering and model tuning techniques.

Projects

Searching Over Enterprise Data

Feb 2024 – present

- Developed a **question-answering system** leveraging NLP techniques and Generative AI models to enable efficient search and retrieval of information from large enterprise document repositories.
- Implemented **embedding and indexing techniques** using Elasticsearch and vector databases, improving search accuracy and speed.
- Technologies: **Mistral Model, Hugging Face, Elasticsearch, Vector Stores, RAG.**

SummarizeIQ

Oct 2023 – present

- Built a **document summarization system** using Generative AI models to streamline information retrieval for enterprises.
- Integrated with Elasticsearch for efficient data indexing and implemented advanced summarization techniques to generate concise summaries from large, complex documents.
- Technologies: **BART Model, Elasticsearch.**

Context-Based Sentiment and Route Prediction

Jul 2023 – Sep 2023

- Designed a **context-aware sentiment prediction model** using Generative AI techniques, providing accurate sentiment analysis by capturing contextual nuances in diverse text inputs.
- Applied **transformer models** (mBERT, BERT-based) to enhance predictive capabilities for various scenarios, improving decision-making processes.
- Technologies: **Transformer Models (mBERT, BERT-based), Kaggle Datasets.**

CareerAlign, (Web Scraping)

Feb 2023 – Apr 2023

- **Description:** Developed a job recommendation system that leveraged LinkedIn data to enhance personalized job matching. Implemented web scraping techniques using Python libraries such as **BeautifulSoup** and **Selenium** to collect relevant job data. Applied **NLP techniques** to extract skills, job roles, and experience requirements, optimizing recommendations for user-specific preferences.
- **Tech Used:** Python, BeautifulSoup, Selenium, spaCy.

GuestPulse, (Natural Language Processing)

Nov 2022 – Jan 2023

- **Description:** Utilized NLP models to analyze hotel reviews and identify attributes that significantly influenced traveler choices. Implemented **sentiment analysis** to gauge customer satisfaction to highlight key features. Developed personalized recommendations based on guest preferences to enhance the user experience.
- **Tech Used:** NLTK, spaCy, Pandas, Python.

- **Description:** Developed a predictive model to identify customer churn in the telecommunications industry. Applied machine learning algorithms such as **Random Forest** and **Gradient Boosting** to analyze customer behavior and predict churn. Employed **feature engineering** and model tuning techniques to enhance prediction accuracy.
- **Tech Used:** Python, Scikit-Learn, XGBoost, Pandas, Matplotlib.

Skills

Machine Learning — Regression, classification, clustering algorithms. | **Natural Language Processing** — Text processing, sentiment analysis, LLMs for question answering. | **Python** — NumPy, Pandas, Scikit-Learn, Statsmodels, Matplotlib, Seaborn. | **Deep Learning:** — Neural networks with TensorFlow, PyTorch, LangChain. | **Elasticsearch & Kibana** — Search and analytics solutions, RAG systems. | **IBM Watsonx** — Watson Discovery, Watson Assistant | **API Development** — Building FastAPI for model deployment.

Certificates

- Data Science Certification from ExcelR
 - Internship in Data Science from AI Variant
- Machine Learning with Python from IBM
 - Masters Program in Data Science from Nasscom

Education

Jamia Hamdard, Masters in Computer Application (MCA)

2020 – 2022 | New Delhi

Jamia Hamdard, Bachelors in Computer Application (BCA)

2017 – 2020 | New Delhi