




# SAQUIB KHAN

Research Assistant

✉ saquibkhan9572@gmail.com    saquib-khan5005    +91 9771826884    Bhubaneswar, India

## PROFESSIONAL SUMMARY

Research intern with results-oriented attitude in AI, NLP, and data analytics. Developed a Spoken Query Retrieval System for FIRE 2024, obtaining relevant documents in English with 85% accuracy and 10% reduction in query processing time. Predicted EPL match outcomes with 75% accuracy using 100,000+ football match records and Power BI dashboards. Comparing 5 dementia prediction ML models, ensemble technique improved accuracy by 92% and reliability by 15%. Also created predictive models for industrial chemical analysis at Tata Steel (83.67% accuracy). Loves using cutting-edge research and computer learning to address challenging, real-world challenges

## RESEARCH EXPERIENCE

### Research Assistant

*FIRE 2024 Conference - Spoken-Query Cross-Lingual Information Retrieval (SqCLIR)*  
[Aug 20, 2024 – Oct 14, 2024]

- Collaborated with a team of 4, including an NLP Engineer, to develop a Spoken Query Retrieval System that achieved 85% accuracy in retrieving relevant documents for monolingual queries in English, Gujarati, Hindi, and Bengali.
- Processed over 5,000 spoken queries using FastWhisper for audio transcription and sentenceTransformers/OpenAI embeddings for semantic similarity analysis.
- Built a retrieval system capable of handling 4 languages, contributing to a novel shared task addressing under-resourced Indic languages at FIRE 2024.
- Reduced query processing time by 30% by optimizing embedding generation and indexing techniques.

### Research Assistant

*Football Analytics Project - English Premier League (EPL) Analysis* [Sep 1, 2024 – Nov 30, 2024]

- Analyzed 20 years of historical football data (2004–2024), processing over 100,000 match records to identify trends in team performance, player stats, and match outcomes.
- Created interactive Power BI dashboards summarizing key insights, which included 50+ visualizations such as win rates, goal distributions, and player performance metrics.
- Identified patterns in team performance, enabling the prediction of match outcomes with 75% accuracy using statistical models.
- Worked in a team of 6 to deliver actionable insights, reducing manual analysis time by 40% through automation in Python.

### Research Intern

*IEEE Access 25' - Comparison of Machine Learning Models for Dementia Prediction*  
[Feb 1, 2025 – Feb 28, 2025](In Progress)

- Compared 5 machine learning models (e.g., Random Forest, SVM, Neural Networks) on a dataset of 1,000 patient records, achieving a peak accuracy of 92% with an ensemble approach.
- Evaluated model performance using metrics such as accuracy, precision, recall, and F1-score, improving prediction reliability by 15% compared to baseline models.
- Processed and cleaned 20+ features from raw patient data, reducing noise and improving model interpretability.
- Documented findings in a manuscript submitted to IEEE Access 25', with contributions from a team of 4 researchers.

## WORK EXPERIENCE

- Tata Steel Private Ltd. | Machine Learning Intern | Python

MAY 2024- Present

  - During industrial training in steel manufacturing, I developed a **Gradient Boosting Regressor ML** model to predict Sulphur, Magnesium, and Lime levels, achieving 83.67% accuracy and 72.24% within  $\pm 0.002$  threshold.
- 360 digiTMG | Associated with INNODATATICS | Machine Learning Intern | Python

Optimizing Machine Downtime using ML | Accuracy: 79%

May 2024 – June 2024

  - Skills:** EDA, Machine Learning and Statistical Analysis, Predictive Maintenance, Programming and Software Development-Proficiency in Python Using libraries and frameworks such as Pandas, Scikit-Learn, Project Management, Collaboration and Communication, CRISP - ML(Q), PostgreSQL
- EY Global Delivery Services | Advance Data Analytics Intern | PowerBI

Flipkart Business Monthly Data Analysis

Dec 2023 - Jan 2024

  - Developed a comprehensive business data analysis dashboard for Flipkart, with a focus on sales performance & customer behavior to provide actionable insights and improve business strategies.

## TECHNICAL SKILLS

- Programming Languages:** Python, Java, MySQL
- Technologies/framework:** PowerBI, MS Excel, ChatGPT, GitHub, VS Code, Google Colab, Jupyter Notebook, Matplotlib, Scikit-Learn, Pandas, Numpy, Plotly, PostgreSQL, Latex, Knime, LLMs

## EDUCATION

- Bachelor of Technology in Artificial Intelligence & Data Science

C. V. Raman Global University | CGPA: 8.54

October 2022 – May 2026

Bhubaneswar, Odisha

**Relevant Coursework:** Machine Learning, Data Mining, Data Structures & Algorithms, Database Management System, Object-Oriented Programming, Computer Network, Computer Organization, Java, Python
- Upper & Lower Secondary Education

Loyola School - **ISC:** 86% in Pure Science | **ISCE:** 93%

March 2009 - July 2022

Jamshedpur, Jharkhand

## ADDITIONAL INFORMATION

- Building Generative AI apps using Ollama and LLMs
- Research Intern in a startup to research for market trends and competitors.
- Actively participating in Webinars and workshops based on ML, DL, NLP & Computer Vision