

# Aadiy Khan

+91-9140965873 | aadiy.23bce10969@vitbhopal.ac.in | <https://www.linkedin.com/in/aadiykhan>  
<https://github.com/AadiyKhan> | Bhopal, Madhya Pradesh, India

## EDUCATION

Vellore Institute of Technology, Bhopal

2023 – 2027

Bachelor of Technology in Computer Science and Engineering

CGPA: 8.38/10.0

## PROJECTS

**FlowBERT:Customer Service Router** | *Python, TensorFlow, Keras, DistilBERT* **Apr 2025 -May 2025**

- Developed end-to-end AI routing system utilizing DistilBERT-based Transformer model to automate ticket categorization and priority assignment, reducing manual triage time by 75%
- Architected Multi-Task Learning (MTL) model in TensorFlow/Keras using custom Lambda layers to simultaneously predict Department Route and Urgency Level from single text input
- Executed data augmentation strategies scaling training dataset from 50 to 600+ records, resolving class imbalance issues and improving model robustness for edge-case customer queries
- Deployed real-time interactive dashboard using Streamlit, enabling support teams to visualize AI-driven classifications and confidence scores for incoming service requests

**Drug Repurposing Pipeline** | *RDKit, Doc2Vec, Gensim, Scikit-learn, NLTK* **Oct 2025 -Dec 2025**

- Architected 5-phase drug repurposing framework integrating chemical informatics and NLP to identify novel therapeutic uses for FDA-approved compounds
- Engineered multi-modal feature extraction system using RDKit for 2048-bit Morgan Fingerprints and Gensim (Doc2Vec) for 100-dimensional clinical embeddings, capturing molecular structure and medical intent
- Developed high-throughput data fusion pipeline synthesizing comprehensive dataset of 2,148 features, optimizing model's ability to recognize complex drug-disease patterns
- Implemented Direct Disease Mapping engine linking DrugBank candidates to clinical indications, utilizing Target Enrichment Analysis to rank and prioritize top-tier candidates for virtual screening
- Automated data cleaning and preprocessing (Regex, NLTK) for thousands of drug records, ensuring high data integrity for large-scale similarity-based medical analysis

**Data Analyst Agent** | *Python, Flask, Pandas, Google Gemini API, Matplotlib, Seaborn*

- Engineered full-stack data analytics platform integrating Google Gemini Pro API for intelligent data insights, enabling natural language queries on uploaded datasets with automated visualization generation
- Implemented comprehensive data preprocessing pipeline supporting CSV, Excel, and text files, featuring automated missing value handling, outlier detection using IQR method, and duplicate removal capabilities
- Built interactive visualization engine generating custom plots (Bar, Line, Scatter, Histogram, Box) with dynamic column selection, reducing manual data exploration time by 65%
- Deployed production-ready web application on Render with real-time AI chat interface, processing user queries and delivering statistical analysis interpretations through RESTful API architecture

## TECHNICAL SKILLS

**Programming Languages:** Python, C++, Java, SQL, JavaScript, HTML, CSS

**Frameworks & Libraries:** TensorFlow, Keras, Transformers, Scikit-learn, Flask, Pandas, NumPy, Streamlit

**Tools & Technologies:** Git, Docker, RDKit, RESTful APIs, Tableau, NLTK, Gensim

**Core Competencies:** Machine Learning, Deep Learning, NLP, Data Structures & Algorithms, Problem Solving

## CERTIFICATIONS & ACHIEVEMENTS

- IBM AI Engineer Professional Certificate, NPTEL Introduction to Machine Learning
- Health Hack Hackathon: Developed real-time AI routing system using DistilBERT achieving 75% reduction in manual processing time
- LeetCode: Solved 350+ algorithmic problems demonstrating strong problem-solving capabilities
- Successfully implemented "Attention is All You Need" research paper from scratch

## ADDITIONAL

**Languages:** Fluent in Hindi, English, and Urdu

**Interests:** Competitive Programming, Open Source Contributions, Badminton, Football