

Hemant Dubey

+91-9910207492

hdubey_be22@thapar.edu

LinkedIn

GitHub

LeetCode

Codolio

EDUCATION

• B.Tech in Electronics and Computer Engineering

2022–Present

Thapar Institute of Engineering and Technology, Patiala, Punjab, India

CGPA: 8.22

• Secondary & Sr. Secondary Education

2019–2021

Modern Vidya Niketan, Sector-17, Faridabad, India

Result: XII: 95.6% | X: 97.8%

WORK EXPERIENCE

• Solid State Physics Laboratory (DRDO) MindFIT | Live

June 2025 – July 2025

ML Project Trainee

- Engineered an automated data pipeline using **Google Apps Script** to collect and augment 200+ psychological test responses, saving Psychologists **2-3 Hours per week** in manual effort.
- Spearheaded the automation of psychological screening by developing the **MindFit Questionnaire**, leveraging the **PHQ-9 dataset** to generate model-driven insights and interactive **dashboard visualizations**.
- Delivered the project in a fast-paced R&D setting; recognized as a **top 1% contributor** on GitHub.
- Secured **Software Management Engineer** role at **TechEagles** based on Intern performance.

PROJECTS

• Movie Recommender | Live

Jan 2024 – Feb 2024

Hybrid model using collaborative and content-based filtering

- Built a hybrid Movie Recommender using **collaborative filtering**, **content-based methods**, and **scikit-learn**, **TensorFlow** on a dataset of 50K ratings from 1K users on 1K movies.
- Achieved **RMSE of 0.76** via Matrix Factorization, KNN; reached **87% accuracy** in genre classification using Logistic Regression. Improved cold-start recommendations by **40%** using cosine similarity.

• Speech Intent Recognizer | Live

March 2025 – May 2025

End-to-end voice command classification from raw audio

- Achieved **98.84% accuracy** on raw audio intent recognition, outperforming transcription-dependent commercial systems. Showcased Data using Seaborn & Matplotlib.
- Enabled **real-time inference** with fine-tuned Wav2Vec2, overcoming latency in **end-to-end** speech models.
- Handled **31 distinct intents** with near-0 misclassifications, surpassing generic voice assistants in precision.

• YouNotes | Live

Aug 2025 – Present

Web app that generates study guides and quizzes from YouTube videos and PDFs

- Built a **content extraction** system using **YouTubeTranscriptAPI** and **PyPDF2** to process video transcripts.
- Designed a **DAG-based AI pipeline** using **LangGraph** to orchestrate summarization, study guide generation, and MCQ creation through specialized agent nodes. Hosted on **Heroku**.
- Integrated as a Subsidiary tool within our **Capstone** project **Thapar Virtual Labs**—a **3-tier Architecture** for hosting assignments and labs online with Quizzes, Manuals, and Experiment Simulations in real time.

POSITION OF RESPONSIBILITY

– **TechEagles-SME**: Freelancing Community Founding Team-Engineer, Providing SaaS

Sep 2025-Present

– **Executive Board Member**, Fine Arts & Photography Society

Oct 2023- May 2025

- *Directed the creative vision for a flagship Art & Music Festival **KaleidoScope** with **5,000+** attendees.
- *Managed the Core team on ideation, execution, and Art Exhibits in just about a **month-long** phase.
- *Concluded tenure as **Head of Sketching** in May 2025, having **led a team** of 30+ artists in the past year.

SCHOLASTIC ACHIEVEMENTS AND AWARDS

- **Amazon ML Summer School 2024&25**: Selected **twice** being 1 out of 80,000+ applicants.
- **Robust-framework-to-detect-Credit-Card-Fraud-Detection-Research-Paper-IEEE(Co-authored)** | May-2025
- **SSB Interview**: Appeared 4 times, Recommended once, reached Conference Interview 2 times.
- **SIP Abacus(Fast Calculations)**: Level-5 2nd Runner Up All-India Competition.
- **CompTIA Network+ | CompTIA | [2022]**

TECHNICAL SKILLS AND INTERESTS

- **Programming Languages**: C++, Python, SQL, HTML/CSS, JavaScript
- **Data Analysis & Interests**: Predictive Modeling, Statistics for ML, Dimensionality Reduction (PCA), pandas, NumPy, scikit-learn, TensorFlow, PyTorch, Keras, Ladder Logic(PLC), AutoCAD(2D)
- **Data Engineering Tools**: MySQL, UV, Delta lake, Databricks, Apache Pyspark, GCP Cloud, MS Excel
- **Tools & Platform**: GitHub, Figma, MATLAB, Streamlit, Google Colab, Jupyter NB, SolidWorks, Linux