

K Soveet Kumar Prusty

+91 7208221665 | soveet.prusty@gmail.com | linkedin.com/in/soveet-prusty | github.com/sovopr

SUMMARY

AI-focused **B.Tech CSE (AI)** student who builds things that sit at the intersection of **systems thinking** and **applied ML** from enterprise-grade cloud data architectures to **low-resource language models** and **graph-based** fraud detection. Comfortable moving between research rigour and engineering pragmatism; have shipped **production systems**, **co-authored papers**, and **filed & published patents** across the same academic calendar.

EDUCATION

Manipal Institute of Technology, Bangalore

BTech CSE (AI) | CGPA: 8.83/10.0

Bangalore, India

July 2023 – July 2027 (Expected)

EXPERIENCE

Research Associate

Max Planck Institute for Human Development (MPIB) & TAPMI Partner Group

Manipal / Berlin

Oct 2025 – Present

- **Lead Engineer** for “Risk Chat Arena”: Architected a scalable **Streamlit** backend handling **50+** concurrent threads to benchmark **Generative AI** risk communication via parallel inference across **6** Frontier Models.
- Built a specialized **RAG** pipeline utilizing **Vector Embeddings** to parse unstructured financial factsheets with **90%** semantic retrieval accuracy.
- Implemented a blind **A/B testing** interface allowing experts to generate evaluation data points, automating logging via **Pandas** to reduce data aggregation time by **40%**.

Technology Consulting Intern

Ernst & Young (EY)

Kolkata, India

June 2026 – July 2026

- Designed end-to-end **GCP × Snowflake** enterprise data lakehouse architectures, producing **SOWs** and technical blueprints covering ingestion (Datastream, Pub/Sub), transformation (Dataflow, Dataproc), and serving (BigQuery, Snowpark) layers for client engagements.
- Architected **CDC pipelines** using **PySpark** on **Databricks** with Bronze-Silver-Gold medallion patterns; integrated **Snowflake** Dynamic Tables and Snowpark ML for downstream model serving.
- Mapped multi-source ingestion flows (CRM, CMS, Mobile, Partner APIs) via Cloud Interconnect and Snowpipe Streaming into client-facing presales architecture diagrams used in active engagements.

Research Intern

Indian Institute of Management (IIM), Calcutta

Kolkata, India

May 2025 – July 2025

- Analyzed market dynamics across **100+ stocks** by processing diverse historical datasets from Yahoo Finance, NSE, and BSE.
- Engineered Python pipelines to ingest, standardize, and process over **1 million** data rows from multiple financial data sources.
- Automated abnormal return calculations, identifying **20%+** anomalous movements in backtests to flag potential misinformation/disinformation.

Automation Intern

Innomotics (formerly Siemens)

Mumbai/Remote, India

June 2025 – July 2025

- Resolved data processing inefficiencies across **100+ legacy spreadsheets**, saving up to **8+ hours** per week.
- Automated data cleaning, merging, and migration from outdated formats using Python (pandas, openpyxl), reducing manual handling by **80%**.
- Developed a Real-Time Predictive Maintenance Dashboard (Flask, Chart.js) for cement kiln sensors, forecasting anomalies **7 steps ahead** via Linear Regression.

Technical Intern

Birla Corp, MP Birla Cement

Kolkata, India

June 2024 – July 2024

- Identified and proposed data-driven strategies that increased sustainability and improved **operational efficiency** by **15%**.
- Researched and presented AI/IoT solutions to senior management; analyzed **10,000+ sales data rows** using Power BI.
- Authored a strategic sustainability report adopted for internal planning and delivered **5+ dashboards** revealing key business trends.

RESEARCH AND PUBLICATIONS

- **Patents (Published):** Capacitance Level Sensing Device (202441053585); Device For Liquid Level Measurement (202441053584); Capacitance Based Liquid Measurement System (202541074957); Capacitive Sensor w/ Squirrel Electrode (202541096119); Capacitance Level Sensing System for Precise Detection (202541061629); Capacitive Liquid-Level Measurement Device (202641050275). **4 additional** under filing.
- Two papers **accepted** at **NQCOMP 2026**: “A Statistically Robust Analysis of the N-Player Quantum Public Goods Game on a Calibrated Noise Model” and “On Representation Error in Quantum Option Pricing: A Diagnostic Study of the Standard Qiskit Implementation.”
- A paper **accepted** at **WCAIAA 2026**: “Performance and Noise Resilience of Grover’s Search Algorithm for the One-Hot Satisfiability Problem.”
- A paper **accepted** at **icSoftComp2025**: “AI-Powered Dashboard for User-Centric Financial Data Analytics and Intelligent Stock Prediction.”

PROJECTS

SovoGPT: Compound AI System for Odia | *PyTorch, Transformers, Mixture-of-Agents* | [GitHub](#)

- Engineered a hierarchical **Mixture-of-Agents (MoA)** architecture orchestrating **3 specialized LLMs**, improving reasoning accuracy on translated benchmarks by **~18%**.
- Built a **Synthetic Data Pipeline** generating **50,000+ high-quality instruction pairs** to solve the Low-Resource “Data Desert” problem for Odia.
- Trained a custom **BPE Tokenizer** optimized for Odia script, reducing token fertility by **37%** vs. standard Llama models and increasing effective context by **1.5x**.

Smurfing Hunter – Blockchain AML Intelligence | *PyTorch Geometric, Flask, D3.js, GNN* | [GitHub](#)

- Engineered a **Graph Neural Network** (3-layer GCN) to classify wallet behaviors, detecting **3 complex money laundering typologies** (Smurfing, Peeling Chains, Mules) across Ethereum-style transaction networks.
- Constructed an advanced **GNN engine** utilizing a **1:30 Weighted Loss** function to reliably process highly imbalanced transaction datasets, enabling robust detection despite sparse fraud signal.
- Built a **D3.js Forensic Dashboard** with interactive Sankey diagrams and automated **SAR generation**, flagging **50+ high-risk anomalies** per batch with real-time risk scoring.

FinZen – Personal Finance Tracker | *Streamlit, MongoDB, scikit-learn, Plotly* | [GitHub](#)

- Created a full-stack app (Streamlit, MongoDB) with a scikit-learn model for AI-powered categorization across **15+ expense types**, achieving **90%+ accuracy** and saving **6+ hours/week**.
- Implemented real-time overspending alerts and interactive Plotly dashboards for personal financial trend analysis.

SKILLS

Programming: C, Java, Python, Swift

ML & AI: PyTorch, Scikit-Learn, TensorFlow

Data Tools: NumPy, Pandas, Matplotlib

Databases: MongoDB, MySQL

Design & Tools: Canva, Adobe Illustrator, Figma, PowerPoint

Methodologies: Statistical Analysis, Data analysis, Deep Learning

Soft Skills: Time Management, Problem Solving, Analytical Thinking, Team Management, Leadership

ACHIEVEMENTS & CERTIFICATIONS

- **Winner, Investigar-MIT (Manipal Techfest):** Secured 1st place in the Electrical and Electronics domain for a research paper presentation.
- **Winner, Turinger (ACM MIT-B):** Won the flagship technical event hosted by the ACM MIT, Bangalore student chapter.
- **Winner, Bid and Pitch (IGNITE MIT-B):** Secured first place in a club-level negotiation and pitch competition.
- **AWS Certified Solutions Architect: SAA-C03** | **AWS Certified Cloud Practitioner: CLF-C02**
- **Quantum Computing Specialisation: Packt** | **Pre-MBA Statistics: IIM Ahmedabad** | **Linear algebra for ML and Data Science: Deeplearning.ai**