

Professional Summary

Data Science professional with **1+ year** experience in machine learning, data analytics, and AI data annotation. Skilled in Python, SQL, and ML frameworks for building analytical models and extracting insights from large datasets.

Education

Master of Mathematics, Indian Institute of Technology Madras	7.43 CGPA (2023-2025)
Bachelor of Science (Mathematics), Patliputra University, Patna	62.2% (2018-2021)
Senior Secondary, Bihar School Examination Board	68% (2016-2018)
Secondary, Central Board of Secondary Education	9.4 CGPA (2015-2016)

Experience

Freelance Data Annotator, Innodata (Remote) Dec 2025 – Present

- Annotate and label **text and image datasets** following structured guidelines to support AI/ML model training.
- Perform **data tagging, classification, and categorization** while maintaining high accuracy and consistency.
- Review annotated data for **quality assurance**, identify inconsistencies, and meet defined productivity targets.

Data Analyst Intern, ERNET India, IIT Madras Research Park July 2024 – June 2025

- Analyzed **FSOC** Communication over a **1.12 km** link in Kohima using Python, processing **180+** days of weather and signal data to detect **outage** causes, quantify downtime and improve network resilience.
- Built **10+** Python visualization tools, **KPI** dashboards and automated reporting scripts to monitor **FSOC** link performance, categorize outage, and enhance **internet** services in remote hilly regions.
- Developed and validated **5+** forecasting models (**RFR, ANN, GRU, LSTM**) to assess performance and reliability, showcasing strong capabilities in model development, evaluation, and ongoing monitoring relevant to risk analytics.

Winter Intern – Applications of ML Techniques, E&ICT Academy, IIT Guwahati Dec 2024 – Jan 2025

- Designed and tuned **5+** deep learning models in TensorFlow/Keras, applying metric-based evaluation (**6+ KPIs**) aligned with model governance and performance monitoring requirements.
- Created **10+** analytical reports, visualizations, and performance summaries in Python to clearly communicate model behaviour, improvements, and data-driven recommendations to mentors and project teams.

Projects

Multilingual Language Classifier: Developed an end-to-end **MLOps** pipeline for classifying **27** Roman-script languages and continent groups on **10,000+** samples using character-level TF-IDF, TruncatedSVD, and Naive Bayes/LDA/QDA models, achieving **96.8%** language and **89.4%** continent accuracy; deployed via **Streamlit** App. [\[verify\]](#)

Music Recommender System: Built and deploy a content-based music recommender by preprocessing **1M+** song lyrics using **NLTK**, leveraging TF-IDF (**5,000** features) and cosine similarity, and delivers real-time **top-5** song recommendations via an interactive **Streamlit** app with Spotify API integration. [\[verify\]](#)

Sequence Modeling for Language Decoding: Built and optimized Seq2Seq transliteration models (**RNN, LSTM, GRU**) for Roman-to-Devanagari on **10,000+** samples, boosting validation accuracy by **8%** through attention mechanisms, beam search decoding and Weights & Biases Bayesian hyperparameter sweeps. [\[verify\]](#)

Multi-label EHR Classification (ML Challenge): Built a deep learning model (**4** dense layers, LeakyReLU, Dropout) to perform multi-label classification on **200K+** EHR embeddings, achieving a Micro F_2 score of **0.8183** on validation data and mastering threshold tuning, data engineering, and evaluation techniques. [\[verify\]](#)

Skills

Technologies & Tools: Python, SQL, NumPy, Pandas, Power BI, Tableau, MS Excel, Git, Streamlit.

Domains: Machine Learning, Deep Learning, NLP, Statistical Analysis, Documentation & Reporting.

Relevant Coursework

Core Courses: Foundation of Machine Learning, Introduction to Deep Learning, Data Structures and Algorithm, Data Analytics Laboratory, Linear Algebra, Probability, Applied Statistics, Numerical Optimization, Multivariable Calculus

Online Courses: OCI Data Science Professional (Oracle), **Data Science** Job Simulation (Lloyds Banking Group), **Data Analytics** (Deloitte), Intro to **AI** (IITM Summer School), Master in **Data Science** (Udemy), Communication Skills (TCS iON), Advance Diploma in Computer Applications (MICT Wizard-Tech), Power BI Workshop.

Leadership, Activities, and Achievements

Scholastic: Secured AIR **417** in IIT-JAM (Mathematics); Qualified **JEE** (2018, 2019) and **GATE** (Data Science, 2025).

Leadership: Volunteer at **Forays 2024** (Math Fest, IIT Madras) managing event operations and participant coordination.

Sports: Represented IIT Madras hostel teams in **Cricket & Football**, achieved 4th position in cricket tournaments and demonstrated teamwork, adaptability, and continuous improvement.