NAMAN AGRAWAL

Data Analyst adept at turning complex data into actionable insights to drive strategic decision-making

CONTACT

- ✉ dr.namanagrawal@gmail.com
- 📞 +91 8439410326
- 📍 Jaipur, India
- Naman-Agrawal19
- Naman Agrawal

TECHNICAL SKILLS

Programming Languages

- **Python, R, Scala**
- **SQL**: PostgreSQL, MySQL CTEs, Joins, Subqueries

Data Science, Machine Learning

- **Data Analysis**:
  PySpark, Pandas, NumPy
- **Machine Learning**:
  Scikit-learn, XGBoost, Random Forests, SVM, Decision Trees, Linear Logistic Regression, KNN
- **Deep Learning**:
  TensorFlow, Keras, Neural Networks, CNN, RNN, LSTM, Transformers
- **Statistics**:
  Hypothesis Testing, T-Test, ANOVA, Chi2 Test, Correlation, Confidence Intervals
- **NLP**:
  Tokenization, Embeddings (Word2Vec, GloVe), TF-IDF
- **Model Deployment**: FastAPI
- **Data Visualization**:
  Matplotlib, Plotly, Seaborn
- **Unsupervised Learning**:
  Clustering (K-means, DBScan), Dimensionality Reduction (PCA)

MLOps & Tools

- **Version Control**: Git, GitHub
- **MLOps**: Docker, MLflow, DVC Kubernetes
- **Cloud**: AWS (EC2, S3)

Productivity

- **MS Office, LaTeX, Markdown**
- **Linux, Windows**

PROJECTS

- **YouTube Comment Sentiment Analysis** 🎧 *(Python, transformer)*
  - **89% accuracy** achieved by Boosting pretrained model through custom preprocessing Hinglish text.
  - **Designed text cleaning pipeline** handling null values, emojis, URLs, code-mixing, and Indian English dialects.
  - **(0.91 F1-score)** Validated on 500 manually labeled comments.
  - **Interactive dashboard** deployed for processing 700+ comments/min.
- **Credit Default Risk Prediction** 🎧 *(Python, CatBoost, Streamlit)*
  - **0.88 F1-score** achieved by developing ML model (CatBoost) on imbalanced data (40k+ rows, 60+ features)
  - **Feature Engineering** by filling nan values, VIF (>6), chi-square, and t-tests, improving precision by 15% versus baseline.
  - **SMOTE** Implementation for class balancing and **hyperparameter tuning** (Bayesian Optimization)
  - **Interactive front-end** deployed for categorizing risk into 4 buckets.
- **Real-Time Payment Fraud Detection System** 🎧 *(XGBoost)*
  - **0.92 F1-score** achieved using XGBoost on 6 million row dataset.)
  - **Solved severe class imabalance** by using SMOTETomek and random undersampling hybrid techniques.
  - **Engineered features** that improved precision by 18%.
  - **Automated Preprocessing and training Pipeline** with (OneHotEncoding, StandardScaler etc.) via scikit-learn Pipeline.
- **Vehicle Insurance Claim Prediction – MLOps Project**
  - **Deployed end-to-end ML pipeline** for predicting insurance claims with **Recall: 1.0, F1 Score: 0.93, and Accuracy: 0.88.**
  - Model Deployment using **FastAPI on EC2 via Docker** with CI/CD using GitHub Actions and ECR.
  - Implemented from data ingestion to model registry on AWS S3.
  - Enabled **real-time predictions** with a FastAPI and used MongoDB Atlas for data storage.

🎓 EDUCATION

📅 Aug 2024 - July 2026 📍 Chandigarh University • Major: Data Science	Master of Science
📅 Aug 2015 - May 2019 📍 Uttar Pradesh Technical University • Major: Mechanical Engineering	Bachelor of Technology

⚙ WORK HISTORY (5+ YEARS)

📅 Sep 2022 - Present (3 years) 📍 Samatrix Consulting Pvt. Ltd.	Data Analyst | Team Lead

- **Leading a 6-member team** delivering data science tools and training to 800+ students, achieving **4.8/5** satisfaction scores.
- **Deploy automated ML pipelines** for detecting real-time fraudulent transactions, improving accuracy by **15%**.
- **Developed Recommender system** by analysing student feedback comments and filled forms, **reducing attrition rates by 20%.**

📅 **Feb 2021 - Sep 2022 (1.5 years)**

📍 **MAT Holdings Inc.- Sonipat**

Mechanical Engineer

- **Automated data workflows** by developing **VBA scripts** to transfer and categorize production data across **Excel** sheets, reducing 70% manual efforts and **saving 10+ hours/day.**
- **Developing Pad Assembly and part drawings**, Printing drawings on CAD as per customer data for PPAP approval as well as Production.
- **Managing complaints and issues** from the production line.

📅 **Aug 2020 - Jan 2021 (6 months)**

📍 **DB Engineering Pvt. Ltd.- Noida**

Mechanical Design Engineer

- Maintain up-to-date documentation of all company processes.
- Develop the Turning CAD Drawing of parts e.g., Knives, Holders, Cutters, Blades etc.
- Make packaging part drawings from OEM for Pad e.g., Shim drawings, Wire Indicators, etc.
- Creating and managing Job Cards according to Production Unit.

📅 **Jul 2019 - Aug 2020 (1 year)**

📍 **Preet Machines Ltd.- Ghaziabad**

Asst. Design Engineer

- Maintain up-to-date documentation of all company processes.
- Develop the detail drawings of Rolling Mill equipments e.g., Roller Table, Pinch Roll, Chain Transfer etc. using AutoCAD.
- Create Erection drawings of General Arrangement as per Layout Bill of Materials(BOM) for purchase
- Collect and analyze the Layout and carry out calculations for the preparation of Drawings.

WRITING & PUBLICATIONS

- **Recurrent Neural Network**: Let the machine remember
- **Optimizers**: When Back Propagation meets brains
- **Logistic Regression**: When Linear is not enough to make the cut
- **Back Propagation**: Pulled Back Harder to Strike Even Stronger
- **Gradient Descent**: No Mercy for High Loss
- **Linear Regression**: Because guessing is not good enough.

ACHIEVEMENTS

- Secured **Process Automation** | Developed VBA scripts to automate cross-Excel data transfer and categorization, reducing manual errors by 70% and saving 10+ hours/day.
- **Kaggle Competitions Contributor** with code, dataset, and discussion contributions.