ASHUTOSH SHARMA

ACADEMIC QUALIFICATIONS

Year	Degree /Board	University /Institution	%/CGPA
2025*	Post Graduate Diploma in Business Analytics	IIM Calcutta, IIT Kharagpur, ISI Kolkata	-
2021	B.Tech + M.Tech. Mechanical Engineering	Indian Institute of Technology (BHU) Varanasi	8.22/10
2015	CLASS XII	Delhi Public School Ghaziabad	92.8 %
2013	CLASS X	Children's Academy	10/10
KEY SKILLS/TOOLS	Regression Analysis, Deep Learning, Python, R, Statistical Data Analysis, PySpark, ETL Tools, PostgreSQL		

WORK EXPERIENCE (23 Months)

Axtria India Pvt. Ltd.	Senior Associate	Noida (Jul '21 - Jun '23)

Data Engineering and Cloud Services	■ **Maintained** unified **datahub** for diverse clients, streamlining data access and processing across organization ■ **Retrieved** and **transferred regulatory** and **label management** data from various sources to **Amazon S3** ■ **Designed** and **optimized** 250+ **ETL pipelines** using **Streamsets** and **Azure Data Factory** with **SCD2 logic** ■ **Designed** framework to **automate** pipeline & job creation saving **4 hrs/day** and **processing cost** by **28%↓** ■ **Created** optimized **DAGS** (Directed Acyclic Graphs) in **Apace Airflow** to **reduce** full load run time by **18%↓**
Data Processing and Analysis	■ **Developed** and **enhanced 300+ publications** for new business requirements using **Pyspark & Spark SQL** ■ **Leveraged Athena** and **SQL** to query source & target data on **S3** for **validating** various ETL process outputs ■ **Utilized Python** Scripts to auto-detect changes in **data granularity**, saving a manual effort of **4.5+** hrs/day↓
Leadership	■ **Led** fifth release of project and **collaborated** with various **cross-functional** teams to **resolve client** issues

AWARDS AND ACHIEVEMENTS

Achievements	■ Secured **AIR 2927(out of 156k)** in **JEE Adv. '16** | Secured **AIR 71** (among **top 0.19 percentile**) in **NEST '16**
Certifications	■ Achieved **Azure Data Fundamentals** and **Databricks Certified Machine Learning Associate** Certifications ■ Achieved **Databricks Accredited Generative AI Fundamentals** & **Green Belt Lean Six Sigma** Certifications
Competitions	■ **National Finalist**(Top 9/940) in **IDB Analytics 3.0** Analytics Case Competitions hosted by **IIM Calcutta '23**

ACADEMIC PROJECTS

Brand Positioning through Perceptual Map (NLP)	■ Created **Perceptual Map** of **Hotel chains** using **OpenAi** embeddings, **BERTopic** and **LDA** from **12k+** reviews ■ Selected no. of topics by **coherence score** & Employed **LR** to describe **thematic** patterns across hotel chains ■ Leveraged **Multidimensional Scaling** to yield **actionable insights** while achieving a **stress score** of 0.096
Customer Targeting (Uplift Modeling)	■ Performed **uplift modelling** to identify customers with **high propensity** to buy dresses after receiving email ■ Implemented **Class Transformation**, **Two model** and **Solo Model** approaches with **RF**, **LR** and **LightGBM** ■ Achieved **Qini coefficient** of 0.038 | Performed **Market Funnel Analysis** & recommended target customers
Mechanism of Action Prediction (Classification)	■ Predicted **MoA** for Drugs using dataset with **23k+** datapoints & **800+** features for **faster drug development** ■ Used **PCA**, **Factor Analysis** and **UMAP** for feature engineering | **MLSMOTE** & **MSKF** for imbalanced dataset ■ Implemented **Xgboost**, **Random Forest**, **Logistic Regression** and **NN** and reduced **logloss** value to 0.0157
Electricity Price Prediction (Time Series)	■ Forecasted Electricity Price in **Day-Ahead** market using temperature, input prices as **Exogenous Variables** ■ Performed **ADF** test for stationarity; Analyzed **ACF & PACF** plots & obtained **MAPE** 17.33% for Naive model ■ Improved Base Model performance to 14.4% using **SARIMAX**, 8.9% using **LSTM**, and 8.1% using **Deep RNN**
Used Car Price Prediction (Regression)	■ Predicted used cars prices(**11.8k+** records) & used **iterative imputer**| detected **multicollinearity** with **VIF** ■ Used **PCA**, **Lasso** & **Ridge** regr. | Performed **residual analysis** | Handled **influential pts.** with **dffits statistic** ■ **Adj. R²** 0.77 (↑11.5%)| Analyzed non-linear models: **AdaBoost**, **XGB**, **RF** and enhanced Adj. R² to 0.84 (↑ 9%)

ADDITIONAL PROJECTS

Landing Page Generation	■ Generated **product description based content** for landing page of ecommerce website using **LLM Agents** ■ Encoded text using **SBERT** and performed **vector storage**, **retrieval** and **query matching** using **Langchain** ■ Used **CrewAI** framework to build **3** agents for creating **taglines**, **detailed descriptions** & brief **product** list
Spam Detection Using LLM	■ Enhanced **user experience** by designing **spam SMS detection** system utilizing **GPT-2** large language model ■ **Fine-tuned** base model to improve **AUC** score from 0.46 to 0.93 & further enhanced it to 0.997 using **LORA**
Bank Customer Segmentation	■ Performed Bank Customer Segmentation using **K-Means** on vector embeddings of **45k+** customer records ■ Utilized **ECOD** for detecting outliers; **Elbow** method for no. of clusters & obtained **Silhouette score** of 0.15
Recommendation Engine	■ Built a **collaborative** Movie RecSys with MovieLens100k dataset using **memory** based strategy with **KNN** ■ Used **Bias subtraction** & **bayesian** based **optuna** hyperparameter tuning to improve MSE from 9.72 to 3.25

POSITIONS OF RESPONSIBILITY & EXTRA CURRICULARS

DS Casebook PGDBA	■ Devised solutions for **ecommerce** & **telecom** industry related cases in 2nd version of **PGDBA DS casebook**
Workshop	■ Organized a **workshop** for guiding the young minds in choosing right career options & conducted Math **quiz**

ELECTIVES : NLP, Bayesian Methods, Financial Risk Management	**INTERESTS :** Watching Cricket, Skipping