SOHAM CHATTERJEE

ACADEMIC QUALIFICATIONS

Year	Degree /Board	University /Institution	%/CGPA
2025*	Post Graduate Diploma in Business Analytics	IIM Calcutta, IIT Kharagpur, ISI Kolkata	-
2022	B.E. (Hons.) Electrical Engineering	Jadavpur University	9.21/10
2018	CLASS XII	Apeejay School, Salt Lake, Kolkata	93.33 %
2016	CLASS X	Delhi Public School, Megacity, Kolkata	96.33 %

KEY SKILLS/TOOLS	SAS Programming, Python, SQL, Power BI, Tableau, Anaplan, Statistical Data Analysis, R Programming

WORK EXPERIENCE (12 Months)

exl Service.com (I) Pvt. Ltd.	Business Analyst	Gurugram (Jul '22 - Jul '23)

Key Responsibilities	■ **Data reconciliation** for **Insurance Claims** data for the East Analytics Region comprising **7 APAC countries**. ■ Book-close tasks for the **State Management Reporting Tool** (SMRT) PowerBI model (**US FP&A** region).
Data Reconciliation	■ **Analysed the inconsistencies** between the Data Lake values & market values for about **1 million records**. ■ Addressed **incorrect source-to-target mapping**, **missing values**, and **incorrectly populated** values. ■ **Correctly collated the data** into Data Lake alongside source teams & **classified causes of loss** using **SQL**. ■ Leveraged **SQL** Server to inspect & **minimize** the initial **variances** of about **20%** in the KPIs, **down to <1%**.
Dashboarding & Planning	■ Presented the **comparison of Data Lake** numbers & **market** numbers with dashboards using **MS Power BI**. ■ Analysed **dataflows**, **transformations**, and performed **variance analyses** on the SMRT Power BI model. ■ Leveraged **SAS** queries to pull out fresh data from the server & **resolved all problems** for smooth reporting. ■ Completed **Anaplan** Model Building **L1** Certification, worked on financial model building, US FP&A region.

AWARDS AND ACHIEVEMENTS

Institute Rank	■ Placed in the **top 20%ile** of academic performers among 62 students of 9th Batch of PGDBA (2023-2025).
Academic Achievements	■ Acquired a spot in the **Top 10 Unstoppable College Champions**, **IIM Calcutta**, Unstop Talent Awards'24. ■ **WBJEE 2018** Rank **353 off 106k** test-takers (among **top 0.3%ile**), Rank **15 off 119** students in undergrad. ■ Offered a **full tuition fee waiver** for the **B.Sc. (Honours)** programme in Mathematical Sciences, **CMI 2018**. ■ Secured **AIR 5** in the **M.Tech** in Quality, Reliability & Operations Research (**QROR**) Examination, **ISI Kolkata**.
Case Competitions	■ **2nd** off **5200+** registrations in **Trilytics'23, PGDBA** with Word Wide Technology (**WWT**), won cash ₹**25k**. ■ **2nd** off **1500+** registrations at **Insight Masters, ISB Hyderabad** at Advaita 2023, won a cash prize of ₹**45k**. ■ **2nd** off **627** participants in the **dashboarding** competition E2A, in Nova Exilaro'24, **XLRI Delhi**, won ₹**12k**.

ACADEMIC PROJECTS

Uber Price Prediction (Regression)	■ Predicted Uber Ride Prices on data containing **200k** records using **OLS Regression** as a baseline model. ■ Treated **outliers**, detected **multicollinearity** using **correlation heatmaps** and **VIF** and tackled it using **PCA**. ■ Used **ANOVA** to judge significance of MLR, and **t-test** for feature selection. Validated all model assumptions. ■ Leveraged **GridSearchCV** to tune hyperparameters for **Ridge** & **Lasso** Regression, achieved best R^2 of **0.72**.
Customer Uplift Prediction (Uplift Modelling)	■ Recommended **optimum range** of customers for maximum profit for the Criteo Uplift Data (~**14M records**). ■ Statistically asserted **CATE**, treated class **imbalance** with **Random Under Sampler** for majority resampling. ■ Predicted uplift using **CausalML** with **XGBoost** as base classifier, achieved best **AUUC 0.64** with **S-Learner**.
TMDb Recommender System (Clustering, NLP)	■ Built content-based recommender system using **TMDb** dataset (~**5k movies** & credits), hosted on Streamlit. ■ **One-hot encoded** categorical type, extracted dictionary types with **AST**, **TF-IDF** vectorized textual features. ■ Used K-Means, DBSCAN & Agglomerative clustering, achieved best **silhouette** score of **0.59** with **DBSCAN**. ■ Built graph network for movies (**networkx**), used **Adamic-Adar Index** to recommend from similar clusters.

ADDITIONAL PROJECTS

AIDS Fatality Detection (Classification)	■ Predicted fatality status using the AIDS Clinical Trials Group Study 175 Dataset (~**2k records**, **23 features**). ■ Treated **multicollinearity** using concepts of **partial correlation**, treated **class imbalance** using **SMOTE**. ■ Achieved a minority **recall 0.68** with Logistic Regression, improved to **0.80** using **CatBoost**, ROC-AUC **88%**.
Image Captioning (CNN, LSTM)	■ Employed transfer learning on **ResNet-50** encoder & **LSTM** based decoder, for automatic image-captioning. ■ Used **nucleus** and **temperature sampling** to build vocabulary with **NLTK**, achieved a **ROUGE** score of **0.2**.
Petrol Price Prediction (Time Series)	■ Forecasted the petrol prices for the upcoming **4 weeks**, using ~200 weeks of historical data using **SARIMA**. ■ Used **ADF** test to verify stationarity, **differencing** for trend removal, **ACF** & **PACF** plots to fix SARIMA orders. ■ Verified residual whiteness via **Ljung-Box** statistic, selected the model with **best AIC**, achieved **MAPE 3.5%**.

POSITIONS OF RESPONSIBILITY & EXTRA CURRICULARS

PGDBA External Relations Cell	■ PGDBA **LinkedIn** follower count ↑**200+** within 4 months, **7k** total followers reached, **YouTube** views ↑**3k**. ■ PGDBA **registrations** ↑**43% , Instagram** views ↑**12k**, promoted brand awareness, resolved aspirants' issues

ELECTIVES : **Deep Learning , Product Analytics**	INTERESTS : **Reading,EDM,Football,Anime**