SUBHAV CHAUHAN

ACADEMIC QUALIFICATIONS

Year	Degree /Board	University /Institution	%/CGPA
2025*	Post Graduate Diploma in Business Analytics	IIM Calcutta, IIT Kharagpur, ISI Kolkata	-
2020	M.Tech. Mechanical Engineering	IIT Mandi	9.06/10
2018	B.Tech Mechanical Engineering	College Of Technology, GBPUAT, Pantnagar	7.642/10
2014	CLASS XII	DAV Centenary Public School J Pur Haridwar UK	92.2 %
2012	CLASS X	DAV Centenary Public School J Pur Haridwar UK	10/10

KEY SKILLS/TOOLS Statistical Data Analysis, Predictive Modeling, Critical Thinking, NLP, Algorithms, Python, R, C++, SQL

AWARDS AND ACHIEVEMENTS

Institute Rank
- Ranked among the top **20 % ile**, for academic performance, among 62 students of PGDBA batch-9 (2023-25)

Case Competitions
- Ranked **1**/586 in WSAC'23 (BITS Hyd.) for developing a **RAG** application for Warehouse Stock Management. Implemented **NLQ** interface using Streamlit with **LangChain** (for SQL query generation using OpenAI API)
- Ranked **5**/608 in OP Analytica (IIT Madaras) for demand forecasting and optimizing warehouse location
- Finalist (top **10**/1194) in ProdoMania, a product management and analytics competition (SJMSOM, IITB)

WORK EXPERIENCE (27 Months)

BYJU'S | THINK & LEARN PVT LTD — **Sr Associate- Content Development** — **Gurugram (Aug '20 - Nov '22)**

RESPONSIBILITIES
- Collaborated in a cross-functional team managing content creation workflow and quality of the content
- Led a team of **12**, providing training and development to optimize key responsibilities and enhance efficiency
- Managed **backend data pipelines** for Q&A, videos, and assessments on the TLLMS and Snowflake platforms
- Administered **content deployment and tracking** within the AWS database for the BTLA Revamp project

ACHIEVEMENTS
- Built team's design proficiency, enabling internal execution and a **~15% reduction** in content creation time
- **Boosted user engagement** by **18%** through feedback analysis and introduction of interactive features
- Recognized with the '**Performer of the Month**' and '**Juggernaut**' awards for the consistent performance

ACADEMIC PROJECTS

Used Cars Price Prediction (Regression)
- Employed multiple linear regression (**MLR**) model to analyze 1500+ data points & predict price of used cars
- Validated model assumptions| Tackled **multicollinearity** using **VIF**, **PCR** & **L1/L2** regularization techniques
- Addressed high **influential** and **leverage** points| Improved **adjusted R^2** from 0.604 (baseline MLR) to **0.777**

UPI Txn. Forecast (Time Series)
- Forecasted UPI transaction volume with time series methods on NPCI data after **ADF** and **ACF/PACF** analysis
- Improved **MAPE** from 11% to 7% by implementing **SARIMAX** (Exogenous var.: Day of Month) over SARIMA

Credit Risk Prediction (Finance)
- Predicted the propensity of customer **loan default** for **15 lakh+** applications with more than **400 features**
- Performed data aggregation, pre-processing & feature engineering| Trained **LGBM** and **CatBoost** models
- Created weighted ensemble model using **soft voting classifier**| Achieved an AUC score of 0.88 & **recall** 0.79

Sentiment Classification (NLP)
- Performed sentiment analysis on financial news data using **TF-IDF** and **Word2Vec** after text preprocessing
- Fine-tuned **FinBERT** and used XG-Boost, Random Forest & Logistic Regression| Achieved max 83% precision
- Trained **ANN** over predictions of above models to predict final sentiment class, improving precision to 86%

Retinal Disease Classification (CNN)
- Developed a classification model using **80k+ OCT images** for identifying **retinal diseases** AMD, DME, CNV
- Fine-tuned **EfficientNet_V2_m** model and achieved **0.997 accuracy (0.5% ↑)** & **0.993 macro-avg F1 score**

Caption Generation (CNN & ViT)
- Built image captioning models using: 1) pre-trained **Visual Transformer** 2) custom CNN-LSTM architecture
- Generated Vocabulary & transformed images| Created custom collate function for captions **batch encoding**
- Employed **GPT2 decoder** for ViT encoder| Implemented **cross-attention**| Achieved Rouge-L score of 0.206
- Employed **teacher forcing** for LSTM and lightweight **EfficientNet** for CNN, achieving a Rouge-L score of 0.27

ADDITIONAL PROJECTS

Cust. Segmentation (Clustering)
- Segmented **1M+** ecom. customers using **Pyspark**; performed feature engineering, outlier treatment & **RFM**
- Used Kmeans, Agglomerative, DBScan for clustering into 5 segm.; best **Silhouette score(0.53)** using K-means

Recommender System (RS)
- Developed a **hybrid** movie RS combining **collaborative filtering** with **content-based graph** on 34k movies
- Leveraged pretrained **Glove word embeddings** and **cosine similarity** to link movies with similar storylines
- Implemented collaborative filtering utilizing user-item interaction matrix factorization with **SVD reduction**

POSITIONS OF RESPONSIBILITY & EXTRA CURRICULARS

PGDBA Conclave
- Managed **public relations** & liaised with **3 partner institutes**| Secured **media coverage** for **Trilytics'24**

Others
- Awarded **second** position in **chess** competition in Aagaaz'19, the inter-year sports tournament at IIT Mandi
- **Coordinated** 'Jugat' competition at Colosseum'18 - Tech. Fest of GBPUAT. Secured previous wins, including **1st** place in 'CADventure' (2017) and 'Junk Yard Wars' (2016), and **3rd** place in 'Illustria' (**Robotics**) (2015)

ELECTIVES : Deep Learning, Computational Finance, PRO | **INTERESTS : Chess, TT, Cosmology**