

Academic Qualifications

***ongoing**

Year	Degree	Institute	CPI / %
2021-present	M. Tech, Industrial & Management Engineering	Indian Institute of Technology, Kanpur	9.14*
2015-19	B. Tech, Mechanical Engineering	Govt. College Of Engineering, Amravati (Maharashtra)	7.69
2015	Class XII	Balbhim Arts, Science & Commerce College, Beed (Maharashtra)	74.46 %
2013	Class X	Champavati High School, Beed (Maharashtra)	89.45 %

Professional Experience

Data Science Intern, Mphasis Limited

- Bias: Detection & Mitigation in Text (NLP)** (Jun '22-Jul '22)
 - Objective: Measuring and Mitigating Bias in the word vectors of text.**
 - Executed **Pre-processing** using Spacy, nltk & custom function for **word tokenization, lemmatization**, stop word removal, punctuations removal, symbols, emoticons, etc.
 - Vectorized words** using GloVe, applied **Frequency Features Reduction** technique, **calculated and removed Bias directions** using PCA to generate **Glove De-Biased word vectors**.
 - Used two metrics: **Bias measurement using Scipy.spatial.distance.cosine function** & **'Predicted word' comparison** before & after De-Biasing.
 - Achieved significant bias reduction in input text corpus & Deployed solution** on AWS Marketplace as **"double hard debias"**.
- Image Augmentation (Computer Vision)** (Jun '22-Jul '22)
 - Objective: Synthetic data generation** using Image Augmentation.
 - Performed **Pre-processing**(cropping images) & **Augmentation**(Transformations like RandomRotate90, GaussNoise, RandomBrightnessContrast) on foreground images.
 - Achieved image throughput of 4/sec with annotations in labelme-coco-yolo & **Deployed solution** on AWS Marketplace as **"data augmentation"**.

Key Projects

- Machine Failure Prediction** [Applied Machine Learning] [Academic Project] (Feb '22-Mar '22)
 - Objective: To predict machine failure** based on **dataset** of 10 features, 10000 samples by handling **data imbalance** using **SMOTE oversampling**
 - Performed **Pre-processing**(Label encoding, null values mean imputation) & **EDA**(Visualized pair-hist plots, **Outlier Analysis** using boxplots).
 - Applied **Logistic Regression, Naive Bayes, SVM, KNN, Decision Tree, Random Forest, XGBoost classifiers**.
 - Achieved best Precision, Recall, F1-scores** as **0.99, 0.98, 0.99**, resp. & **Accuracy** of **0.97** with **XGBoost model**.
- Forecasting SBI Stock Price with Time Series Analysis** [Time Series Analysis] [Self Project] (Jul '22-Aug '22)
 - Objective: To forecast SBI stock price for next 45 days** using 1095 data samples of 4.5 years from 01/2017 to 06/2021.
 - Checked for Stationarity, Seasonality, Trend** using **Dicky-Fuller test, stationarised time series by Differencing(d)**.
 - Plotted PACF** (Partial Autocorrelation function) and **ACF** (Autocorrelation function) **to find optimal parameters p, d, q**.
 - Applied Time Series models – **AR, MA, ARMA, ARIMA, SARIMA, Exponential Smoothing**.
 - Predicted stock price for next 45 days**, based on best tuned **SARIMAX model** order=(1,1,1), seasonal order=(1,1,1,4) with **RMSE 46.329**.
- Forecasting Monthly Champagne Sales** [Time Series Analysis] [Self Project] (Jul '22-Aug '22)
 - Objective: To predict 2 years of monthly champagne sales** from past 9 years of sales data using **time series** techniques.
 - Decomposed the time series** into its **components** by **analysing trend, seasonality, noise**, etc.
 - Checked **stationarity** using **ADF-test (Augmented Dickey-Fuller)** and **stationarised time series by Differencing(d)**.
 - Plotted PACF** (Partial Autocorrelation function) and **ACF** (Autocorrelation function) **to find optimal parameters p, d, q**.
 - Predicted champagne sales for next 2 years** by **Applying AR, ARIMA, SARIMA models** & utilised **RMSE and MAPE** as **evaluation metrics**.
- Analysis of life expectancy using Panel Data Regression** [Statistical Modelling for Business Analytics] [Academic Project] (Aug '21-Sep '21)
 - Objective: To analyse life expectancy** using Panel data of 21 unique features & 1288 data observations for 116 countries in decade 2008-2018.
 - Performed **Pre-processing**(Null values filled by **Mean Imputation** technique), calculated **measure of fit(R²)**, **correlation matrix**, **Hausman test** performed for **Endogeneity & Breusch-Pagan test for heteroskedasticity**, checked **multicollinearity** using **VIF** & checked **omitted variable bias**
 - Applied **pooledOLS & panelOLS** using top 7 correlated features using multiple combinations of **time and entity fixed effects**.
 - Feature elimination done using **RFE**(recursive feature elimination) based on **p-value** and **PanelOLS** with **time fixed effect** has **highest R²= 0.4667**.
 - Performed **t-stat Hypothesis testing** & **concluded significant features** affecting Average Life Expectancy for country from **"best panelOLS model with time fixed effect"** as **gdp_capita, grossdomsavings_gdp, exports_gdp** with **p-values 0.0002, -0.0295 & 0.0313**.
- Customer Segmentation Engine: Analysing behaviour & associated Marketing strategies** [Data Mining][Self Project](Aug '21-Sep '21)
 - Objective: To Segment customers into unique clusters** for making effective, specific & optimal marketing strategies.
 - Pre-processed** a dataset for **nulls & duplicates**, studied **descriptive stats**, **visualized** using **pair-plots, data analysis**, checked for **class imbalance**
 - Performed **data visualization** and **plotted elbow graph** to identify parameter-**"Number of Optimal Clusters"** for **K-Means clustering algorithm**
 - Identified 5 unique clusters** of customers based on their spending behaviour based on Age, Annual Income & Spending Score.
- Descriptive analysis of educational quality of post-graduation at IIT Kanpur** [Marketing Research][Academic Project](Feb '22–May '22)
 - Objective: To analyse-yield insights** on MTech program by identifying key features, merits-demerits & help management decide policies-curriculum
 - Formulated **hypothesis**, collected 129 samples in online surveys, focus groups, personal interviews & performed exploratory, descriptive analysis in **SPSS** by convenience random sampling. Analysed data by **statistical tests**(One/two-sample t-test, Chi-squared test) at **90% CL** in hypothesis testing
 - Identified key features as **good infrastructure, rich libraries & latest curriculum** for better **career opportunities** & establishing **IIT brand value**
- Designing of 3 mini-Projects using HTML, SQL & PHP** [Computer Aided Decision Systems] [Academic Project] (Feb '22-May '22)
 - Objective: To design ER diagrams** for 'bank loan application' satisfying **entity relationships & cardinalities** to develop HTML forms & database

- **1.Customer Data Collection Form:** Collected customer data using **HTML** form, stored into **MariaDB** & then sent into **IIT K servers** using **PHP**.
- **2.Billing & invoice Counter for store:** Designed **HTML** form for creating invoice & providing automatic bill invoice using **JavaScript & PHP**.
- **3.Bank Loan Application Portal:** Designed 3 **html** multi-page forms for Bank customer, officer, manager & used **MariaDB** to store & query data.
- Employed **PHP tool for server-side scripting** & linked designed websites to databases and handled IIT K web-server using terminal called **PuTTY**

8. End-to-end Sales Analysis for Electronic Store using Power BI [Microsoft Power BI Dashboard] [Self Project] (Jul'22-Aug'22)

- **Objective:** To create **Power BI dashboard** for insights of sales by product, by market & by segment.
- Performed **EDA**(Cleaned & Transformed data in usable quality) & created attributes deliveryYear, deliveryDays, etc. for extracting hidden insights.
- Utilised visuals, charts and geo map elements in **Power BI dashboard** for extracting insights of sales by product, by market & by segment which would help in **making data driven decisions**.

Key Academic Courses			*ongoing
Applied Machine Learning	Statistical Modelling for Business Analytics	Introduction to Computing	
Data Mining and Knowledge Discovery*	Causal Inference Methods for Business Analytics*	Probability and Statistics	
Marketing Research	Computer Aided Decision Systems	Operations Research for Management	

Technical & Soft Skills

- **Programming Languages and Tools:** Python | SQL | Microsoft Power BI Desktop | MS Office | HTML | PHP, etc.
- **Libraries & Packages:** numpy | pandas | SciKit-learn | scipystats | statsmodels | tensorflow | glove | word2vec | openCV | matplotlib | NLTK | seaborn, etc.
- **Soft Skills:** Leadership | **Team Management** | **Problem Solving** | Communication

Positions of Responsibility

- Member of “**Media N Culture**” team as a “**Treasurer**” (Jul'21-present) & **TA for IME603 course** (Jul'22-present) in **DIME, IIT Kanpur**
- “**Class Representative**” (**CR**) for 1st and 2nd year (**2015-2017**) in BTech. college GCOE, Amravati, Maharashtra.
- “**Student’s Finances Manager**” in BTech. college **Hostel- SATPUDA**, GCOE, Amravati, Maharashtra, for 2 years.
- “**Head of Dance Department**” for 1 year (2018-19) & “**Co-Head of Dance Department**” for 2 years (2016-18), conducted street plays(Nukkad natak) on multiple occasions about **SOCIAL AWARENESS**, **Women Safety Awareness**, etc.

Achievements & Technical Certifications

- Secured **AIR 523 out of 1.2 lac+ students** in GATE 2021 in Mechanical Engineering stream Paper with **99.57 Percentile**.
- Passed **Problem Solving Basic Skills & Python Basic Skills Certification** Exams from **HackerRank** & achieved **5-Star Gold Badge for Python skills**.
- Completed soft-skills training under TCS ion digital learning hub: “TCS iON Career Edge - Young Professional” Certification (2 weeks duration).
- Completed “Python Hands-On 46 Hours, 210 Exercises, 5 Projects, 2 Exams”- An UdeMy Course
- Attended webinar on “Insights on Leveraging the power of AI & ML in Entrepreneurship”