

ACADEMIC DETAILS			
YEAR	QUALIFICATION	EDUCATIONAL INSTITUTION	PERFORMANCE
2019-Pursuing	M.Tech (Industrial & Management Engineering)	Indian Institute of Technology, Kanpur	7.9* (CPI)
2014-18	B.Tech (Mechanical Engineering)	Kalyani Govt. Engg. College, West Bengal	9.18
2014	Class XII (WBCHSE)	Onda High School	91%
2012	Class X (WBBSE)	Onda High School	92.71%

\*upto 3<sup>rd</sup> semester

SCHOLASTIC ACHIEVEMENTS
<ul style="list-style-type: none"> <li>Secured All India Rank <b>416</b> in <b>GATE-2019</b> (Mechanical) amongst <b>167376</b> students, conducted by IIT MADRAS</li> <li>Received scholarship from <b>Indian Oil Corporation Ltd.</b> under <b>Indian Oil Educational Scholarship Scheme-2014</b></li> </ul>

SUMMER INTERNSHIP
<p><b>Data Science Intern at Harvesting</b> (April'20-June'20)</p> <p><b>Objective:</b> To prepare chatbot in multiple languages (English, Hinglish, Hindi)</p> <ul style="list-style-type: none"> <li>Objective of the project was to create a <b>Chatbot</b> to communicate with farmers and deploy it on <b>Whatsapp</b></li> <li>Used feature of <b>Dialogflow</b> to create chatbot like Agent creation, intent, entity, training, integration etc</li> <li>Trained bot by giving every possible user expression and fetch the details of farmers</li> <li>Deployed the Chatbot in Whatsapp using <b>Twilio</b></li> </ul>

ACADEMIC PROJECTS	
Data Mining and Knowledge Discovery	<p><b>Yellow Taxi Demand Prediction in New York City</b> (Sep'19-Nov'19)</p> <ul style="list-style-type: none"> <li>Predicted taxi demand for New York city from the given dataset of 19 features</li> <li>Performed Data visualization, Data preparation using <b>K-Means Clustering</b> and <b>Time Binning</b></li> <li>Applied different models such as simple moving average, weighted moving average, <b>Exponential weighted moving average</b></li> <li>Applied Regression models were <b>Linear Regression, Random forest Regressor</b> and <b>Boosting</b></li> <li>Models were tuned and compared on <b>MAPE</b> metric and <b>MAPE</b> was 0.1293 of <b>Exponential weighted moving average</b></li> </ul>
	<p><b>Telecom Customer Churn Prediction</b> (May'20-June'20)</p> <ul style="list-style-type: none"> <li>The dataset contains 21 features such as "internet", "online security", "paperless billing" etc</li> <li>Applied <b>SMOTE</b> and <b>RFE (Recursive Feature Elimination)</b> to improve the baseline model to predict churn</li> <li><b>Logit</b> and <b>Probit</b> models were used for classifying the churn class</li> <li>Both the models had similar kind of results and best model was <b>Logit</b> model</li> <li>Reported an <b>accuracy</b> of <b>Logit</b> model was about <b>79%</b>, <b>Precision 73.8%</b> and <b>Recall 62.4%</b>, <b>AUC of ROC</b> was <b>0.83</b></li> </ul>
Applied Machine Learning	<p><b>Netflix Movie Recommendation System</b> (Mar'20-April'20)</p> <ul style="list-style-type: none"> <li>The dataset contains MovieID followed by UserID, Rating and Date</li> <li>Performed preliminary data analysis and <b>temporal train-test splitting</b> and created <b>sparse matrix</b> from dataframe</li> <li>From sampled training data created <b>13 initial features</b> for regression such as UAvg, MAvg, GAvG etc</li> <li>Applied different models such as <b>Surprise baseline model, Surprise KNN model, Matrix Factorization</b> techniques</li> <li>Applied XGBoost Regressor with surprise models and got the least <b>RMSE 1.0726</b> from <b>Matrix Factorization</b></li> </ul>
	<p><b>Analysis of the Factors Affecting Sales Price of House in King County, USA</b> (Jan'20-Feb'20)</p> <ul style="list-style-type: none"> <li>Carried out <b>multivariate statistical regression analysis</b> to study the factors influencing house prices with <b>19 features</b></li> <li>Determined correlation matrix and checked for <b>Multicollinearity</b> and performed <b>EDA</b></li> <li><b>Breusch-Pagan</b> test showed there was heteroskedasticity in the data, hence <b>Heteroskedastic robust errors</b> were used</li> <li><b>Adjusted R-squared</b> with and without robust error was 0.694 &amp; 0.676 respectively</li> </ul>
Statistical Modelling for Business Analytics	<p><b>Statistical Analysis on Factors influencing Life Expectancy Panel Dataset</b> (April'20-June'20)</p> <ul style="list-style-type: none"> <li>The dataset consists of <b>193 countries</b> from year 2000-2015 with 22 Features</li> <li>Checked <b>Heterogeneity</b> across countries (or entities) and years</li> <li>Applied Panel Models were <b>Pooled</b> regression, <b>Entity &amp; Time Fixed effects</b> regression and <b>Random Effects</b> regression</li> <li>The highest Adjusted R-squared was 0.9482 with <b>Binary regressor</b> model for entities</li> <li>Conducted <b>Hausman</b> test and test for checking <b>Panel effect</b></li> </ul>

COURSEWORK AND SKILLS	
Relevant Courses	<b>Data Mining and Knowledge Discovery</b>   <b>Statistical Modelling for Business Analytics</b>   <b>Probability &amp; Statistics</b>   <b>Applied Machine Learning</b>   Introduction to Computing   Operations Management   Business Management using Cloud   E- Supply Chain Management   Operations Research for Management
Technical Skills and Certifications	Python (NumPy, Pandas, Matplotlib, Seaborn, Scikit Learn, SciPy)   SQL   MS Excel SQL for Data Science Machine Learning : Hands-on Python In Data Science

POSITIONS OF RESPONSIBILITY
<ul style="list-style-type: none"> <li>Students' senate nominee at <b>Senate Curriculum Development and Monitoring Committee</b> (2019-20), IITK : Addresses the issue related to review and revise of both content and conduct of courses</li> <li>Serving as a teaching assistant to an IME course <b>Statistical Modelling for Business Analytics</b> (Sep'20-Present) : Managed and provided support to 64 students and administered exams</li> <li>Student Member at <b>Indian Society of Heating Refrigerating and Air-Conditioning Engineers</b>, Kolkata Chapter during the year 2016-17</li> </ul>