

Data Science Professional with 4+ years of experience. Currently working in B2B segment with focus on alternative data. Also has experience in US healthcare domain. Proven academic records and demonstrated analytical capabilities on competitive platform.

Work Experience/Internships		
Senior Data Scientist	GlobalData Plc	July 2019 - Present
Patents Company Name Normalisation	<ul style="list-style-type: none"> <li>Created an algorithm to map company name variants with an existing company naming system</li> <li>Utilised text cleaning, patent features like family &amp; inventor details and company features like headquarter, revenue and several user defined metrics for developing the matching algorithm</li> <li>Algorithm was able to cover 95% of all the patent applicants. New variant database can be utilized in various other data sources like news, deals and social for faster information processing</li> </ul>	
Modelling Projects	<ul style="list-style-type: none"> <li><b>Disruptor Innovation Identification:</b> Created an ML model to classify an article into innovation or not. By fine tuning BERT embedding, an accuracy of 93% was achieved. Also built a second stage model to summarise the article by finding benefits and challenges in concerned innovation</li> <li><b>Influencer Recognition:</b> Created a filter identifying individuals using the profile images from twitter for identifying the social media influencers. Model was built using Convolutional Neural Network (CNN) with an accuracy of 95% and reducing manual effort by more than 90%</li> <li><b>Consumer Review Insights:</b> Created an automated cluster and topic generation for driving insights from drug reviews data. Used Hugging Face Transformer for vectorising and density based hierarchical clustering. For topic generation used spacy with zero-shot similarity metric</li> <li><b>Sentiment Model:</b> Built an ML model for measuring the sentiments of tweets using the internal and public data. Gathered 3 million training data from various open sources like universities &amp; data platforms. Utilised the word2vec approach and deep neural net for training the sentiments</li> </ul>	
PR Reports, POCs, Automations	<ul style="list-style-type: none"> <li>Created a report on Vegan based on social data using text and keyword search on the corpus</li> <li>Created a report &amp; interactive dashboard on beers to show case capabilities of utilising blogs data</li> <li>Automation of several legacy one pager reports using python to save manual time and efforts</li> </ul>	
Analytical Dashboards	<ul style="list-style-type: none"> <li>Verdict Media websites traffic monitoring for sales team using Google BigQuery and Data Studio</li> <li>Dashboard using social indicators to analyse the effect of Covid-19 on companies and sectors</li> <li>Competitive Intelligence Dashboard for company performance benchmarking using Plotly dash</li> <li>Created a UI for Analytics Hub dashboard for analysis &amp; metric generation from alternative data</li> </ul>	
Data Scientist	UnitedHealth Group	Oct 2017 – June 2019
Retention Driver Analysis	<ul style="list-style-type: none"> <li>Worked on finding key drivers influencing member retention during annual enrolment period</li> <li>Built logistic, random forest and gradient boosting model for identifying key driving factors</li> <li>Utilising concept of shapley value and partial dependent plots for quantifying the drivers impact</li> </ul>	
CAHPS Project	<ul style="list-style-type: none"> <li>Worked for enhancing NPS &amp; Medicare Plan Star ratings by utilising the member survey data</li> <li>Analysed contract &amp; plans star rating and likelihood to reach 4 star rating using logistic classifier</li> <li>Predicted members who are likely to provide low scores in survey using random forest model</li> </ul>	
Member Tenure Prediction	<ul style="list-style-type: none"> <li>Worked for personalising prospect acquisition, member engagement and retention initiatives</li> <li>Worked on random forest survival model for predicting member churn and its tenure with UHC</li> <li>Performed Customer Life Time Value analysis using the predicted tenure for marketing strategy</li> </ul>	
Internship (6 months)	<ul style="list-style-type: none"> <li>Worked for Provider side of the business to improve experience and reduce cost by digitisation</li> <li>Project involved working on Big Data platforms like Hive &amp; Spark for data collection &amp; analysis</li> <li>Created a complete automation of process from data ingestion to data transformation and update</li> <li>Received <b>Bravos 2 times</b> from the Product Owner &amp; the Manager for the collaborative work done</li> <li>Received PPO to serve as full time data scientist &amp; graded excellent in the Internship evaluation</li> </ul>	
OPTUM - Data Science Challenge	<ul style="list-style-type: none"> <li>1<sup>st</sup> Runner up world-wide (1<sup>st</sup> in India) for wining data science challenge among employees</li> <li>Built a PCP recommendation model that can reduce cost by choosing right PCP for members</li> <li>Final Model was built using 10 variables with lots of advanced feature engineering technique</li> <li>Achieved an AUC of <b>86.66</b> on public leader board using a single model with tuned parameters</li> </ul>	

Software Engineering Associate, IT	Accenture Services Pvt. Ltd.	Jan - May 2016
------------------------------------	------------------------------	----------------

Accessibility Testing	<ul style="list-style-type: none"> <li>Testing role for US-based software client to make web applications accessibility compliant</li> <li>Worked involve preparation of test cases &amp; scripts, running of test scripts, and logging bugs</li> <li>Received appreciations from the manager for the work done &amp; being the <b>Best performer</b></li> </ul>
-----------------------	--

Academics			
Degree/Examination	Year	Board / University	% / CGPA
Post Graduate Diploma in Business Analytics	2016-2018	IIM Calcutta, ISI Kolkata, IIT Kharagpur	8.74
B.Tech, Mechanical Engineering	2011-2015	NIT Patna	8.80
Class XI-XII	2009-2011	CBSE	91.0
Class X	2008-2009	CBSE	90.5

Academic – Data Science Projects		2016-2018
Retrieval based Chat Bots	<ul style="list-style-type: none"><li>Developed an algorithm for retrieval based chat bot on public available Ubuntu Dialogue Corpus</li><li>Applied methods that include Cosine Model-TF-IDF using n-grams, <b>Dual Encoder LSTM</b> (RNN)</li><li>Achieved a recall measure of <b>77%</b> using the Cosine model and a recall of <b>88%</b> using LSTM(RNN)</li></ul>	
Stock Market Movement Prediction using Social Sentiments	<ul style="list-style-type: none"><li>Modelled the social sentiments on stock price to predict the direction for the next trading session</li><li>Extracted features using Information Retrieval based techniques from the top 25 daily news data</li><li>Achieved an accuracy of <b>59.52%</b> on test set using SVM and simulated a buy/sell trading strategy</li></ul>	
Robust Regression using Campbell's Method	<ul style="list-style-type: none"><li>Assigned proper weights to observations of many public datasets by iterative Campbell's method</li><li>Found the outliers using the weights and implemented weighted regression for prediction in R</li><li>Analysed the properties of the Campbell's method of estimation using Monte Carlo simulations</li></ul>	

Analytics Related Achievements		Kaggle	2017 - 2020
Sberbank Russian Housing Market	<ul style="list-style-type: none"> <li><b>Silver Medal:</b> Ranked of <b>56</b> out of 3274 teams for <b>Price Prediction</b> of housings in Russian market</li> <li>Used extensive data cleaning and applied an ensemble learning approach for modelling the price</li> </ul>		
Booz Allen Hamilton's Data Science Bowl	<ul style="list-style-type: none"> <li><b>Bronze Medal:</b> Ranked <b>190</b> out of 1972 teams for <b>Lung cancer prediction</b> from the medical scans</li> <li>Used <b>Xgboost</b> and Neural Network for the purpose of prediction to save life and to reduce cost</li> </ul>		
Quora Question Pairs	<ul style="list-style-type: none"> <li>Ranked <b>402</b> against 3307 teams for modelling <b>Semantic Similarity</b> of question asked on Quora</li> <li>Used LSTM and Xgboost along with feature engineering to find similar questions in the test set</li> </ul>		
Other Competitive Achievements	<ul style="list-style-type: none"> <li>Highest Rank achieved <b>510</b> among 100k (<b>top 1%</b>) data scientist world-wide on platform <b>Kaggle</b></li> <li><b>Silver Medal:</b> <b>88</b> out of 1832 teams on synchronous kernel only competition Instant Gratification</li> <li><b>Bronze Medal:</b> Ranked <b>86</b> against 385 teams in Sea Lion <b>Population Count</b> by NOAA Fisheries</li> <li><b>Bronze Medal:</b> <b>328</b> out of 6381 participants on fraud transaction detection identification by IEEE</li> <li><b>Bronze Medal:</b> <b>240</b> against 3775 teams in US home sales price prediction (Z-estimate) by Zillow</li> </ul>		

Co-curricular and Extra-Curricular Activities
<ul style="list-style-type: none"> <li>Secured <b>1<sup>st</sup></b> runner-up position in StockMart, a live stockmarket event, conducted under Anwasha, IIT – Patna in 2013</li> <li>Secured <b>2<sup>nd</sup></b> runner-up position in Artysticks, an architecture event, conducted under Anwasha in IIT – Patna in 2013</li> <li>Participated in several robotics event &amp; workshops at IIT Guwahati (2012), ISM Dhanbad (2011) &amp; NIT Patna in 2012</li> <li>Ranked in <b>top 2%</b> among 5 lakhs candidates in IIT-JEE in 2012 and <b>top 2%</b> among 11 lakhs candidates in AIEEE 2011</li> <li>Rewarded a certificate of merit for excellent performance in All India Talent Search Examination(AITSE) held in 2003</li> </ul>

Technical Skills and Certifications
<ul style="list-style-type: none"> <li>Programming Languages/Tools - Python, R, SQL, BigData - Hadoop (Hive), Google BigQuery, MS-Excel</li> <li>Data Science tool kits: SciKit-learn, Scipy, numpy, pandas, statsmodel, plotly, NLTK, H2O, Keras, SpaCy, Transformers</li> <li>Certifications: Executive Data Science specialisation- Johns Hopkins University, Machine Learning - University of London, IBM certified Data Science Professional</li> </ul>