








# Sidharth Kumar

[LinkedIn](#) [Github](#) [Kaggle](#) [Credly](#)

Ph: +91 – 9902785537, 9470703519  
Email: Sidharth05kumar@gmail.com

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.

2008-09	2010-11	2011-15	2016	2016-18	2017-19	2019-Present
						
Class X CBSE Score: 90.5 %	Class XI-XII CBSE Score: 91 %	B.Tech - NIT Patna Grade: 8.80	Accenture, Software Associate (5 months)	PGDBA - IIM Calcutta, IIT Kharagpur, ISI Kolkata Grade: 8.74	UnitedHealth Group, Internship & Full time Data Scientist (6+14 months)	GlobalData Plc, Senior Data Scientist (25+ months)

Key Projects		GlobalData Plc	July 2019 - Present		
<ul style="list-style-type: none"><li>▪ <b>Patents Company Name Normalisation:</b> Created an algorithm to map patent assignee name variants with an existing company naming system. Utilised various patent &amp; company features and several textual metrics for developing the matching algorithm. Algorithm was able to cover 95% of all the patent applicants and serves as database for other sources.</li><li>▪ <b>Disruptor Innovation Identification:</b> Created a language model to classify an article into innovation or not. An accuracy of 93% was achieved by fine tuning BERT embedding for classification task. Also built a second stage model to summarise the article.</li><li>▪ <b>Influencer Recognition:</b> Created a model to identify human accounts using the profile images from twitter. Model was built using Convolutional Neural Network (CNN) with an accuracy of 95% and serves as a filter for influencer identification.</li><li>▪ <b>Consumer Review Insights:</b> Created an automated cluster and topic generation for driving insights from drug reviews. Used transformers as an input to density based hierarchical clustering and zero-shot similarity metric for topic selection.</li><li>▪ <b>PR Reports and Automations:</b> Created a report on Vegan based on social data using predefined business text and keyword searches. Created an interactive dashboard on beers to show case capabilities of utilising blogs data. POC on automation of several legacy one pager reports using python to save manual time and efforts.</li><li>▪ <b>Analytical Dashboards:</b> Developed Verdict Media websites traffic monitoring for sales team using Google BigQuery and Data Studio. Created a dashboard using social indicators to analyse the effect of Covid-19 on companies and sectors. Built a Competitive Intelligence Dashboard for company performance benchmarking using Plotly dash framework. Created a UI for Analytics Hub dashboard for analysis &amp; metric generation from alternative data.</li></ul>					
		UnitedHealth Group			
		Oct 2017 - June 2019			
		<ul style="list-style-type: none"><li>▪ <b>Retention Driver Analysis:</b> Worked on finding key drivers influencing member retention during annual enrolment period. Built logistic, random forest and gradient boosting model for identifying key driving factors. Used the concept of shapley value and partial dependent plots for quantifying the impact of key drivers on retention.</li><li>▪ <b>CAHPS Project:</b> Worked for enhancing NPS &amp; Medicare Plan Star ratings by utilising the member survey data. Analysed contract &amp; plans star rating and likelihood to reach 4-star rating using logistic classifier. Predicted members who are likely to provide low scores in survey using random forest model.</li><li>▪ <b>Internship (6 months) -</b> Worked for Provider side of the business to improve experience and reduce cost by digitisation using big data platforms for data analysis. Created a complete automation of ETL process. Received <b>Bravos</b> awards twice and was graded excellent in internship evaluation. Also received pre-placement offer to serve as full time data scientist.</li><li>▪ <b>OPTUM - Data Science Challenge -</b> 1<sup>st</sup> Runner up world-wide (1<sup>st</sup> in India) for wining data science challenge among employees. Built a PCP recommendation model that can reduce cost by choosing right PCP for members. Final model was built using lots of feature engineering and parameter tuning. Achieved an AUC of <b>86.66</b> on private leader board.</li></ul>			
Key Projects		Accenture	Jan 2016 - May 2016		
<ul style="list-style-type: none"><li>▪ Worked for US-based software client to make web applications accessibility compliant. Involved in preparation of test cases &amp; scripts, running of test scripts, and logging bugs. Received appreciations from the manager for the work done.</li></ul>					
Academics Projects		2016 - 2018	Kaggle Projects and Achievements	2017 - 2020	
<ul style="list-style-type: none"><li>• <b>Retrieval based chatbot</b> using Dual Encoder LSTM (RNN) model with a recall of 88% on publicly available Ubuntu Dialogue Corpus</li><li>• <b>Stock Market Movement Prediction</b> using social discussions with an accuracy of 60% without any financial feature in the model</li><li>• <b>Research on Robust Regression</b> for small dataset. Implemented Campbell’s method and simulated using Monte Carlo Simulation</li></ul>			Highest Rank achieved <b>510</b> among 100k ( <b>top 1%</b> ) data scientist on the platform		
			<b>Problem</b>	<b>Rank (Tier)</b>	
			Sberbank Russian Housing Market Price Prediction	<b>56 / 3274 (Silver)</b>	
			Data Science Bowl – Lung Cancer Prediction	<b>190 / 1972 (Bronze)</b>	
			Instant Gratification – Kaggle Internal competition	<b>88 / 1832 (Silver)</b>	
			NOAA Fisheries – Sea Lions Count	<b>86 / 385 (Bronze)</b>	
			Zillow Housing – Zestimate Prediction	<b>240 / 3775 (Bronze)</b>	
			IEEE Fraud Detection in Transactions	<b>328 / 6381 (Bronze)</b>	
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 Anwesha, IIT – Patna in 2013</li><li>▪ Secured <b>2<sup>nd</sup></b> runner-up position in Artysticks, an architecture event, conducted under Anwesha 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>					
Programming		Tools	Data Science & ML/DL Frameworks	Certifications	
<ul style="list-style-type: none"><li>• Python</li><li>• R</li><li>• SQL</li></ul>		<ul style="list-style-type: none"><li>• Hadoop – Hive</li><li>• Google-Big Query</li><li>• Google Data Studio</li></ul>	<ul style="list-style-type: none"><li>Scikit-learn, Scipy, Numpy, Pandas,</li><li>Plotly, Dash, NLTK, H2O,</li><li>Keras, SpaCy, Tranformers</li></ul>	<ul style="list-style-type: none"><li>• IBM Data Science Professional</li><li>• Executive Data Science – JHU</li><li>• Machine Learning – UL</li></ul>	