

Madhur Jain

E-Mail: mdhr.jain@gmail.com

LinkedIn: <https://www.linkedin.com/in/madhur-jain-a86021105/>

Mobile: +91-8793124247

Profile:

IT professional with 6+ years of Industry experience working as Data Scientist using Machine Learning NLP and AI with Python.

Experience Summary:

- Presently working with Larsen and Toubro InfoTech as a Senior Software engineer.
- Master of Technology from Birla Institute of Technology and Science, Pilani with specialization in Data Analytics.
- Having a cognitively engaging full-time IT work-experience as a Data Science Engineer.

Key Skills

- Expertise in Applied Machine Learning, Neural Networks & Deep Learning for NLP tasks
- Experience in using various NLP libraries and techniques like: Spacy, genism, NLTK, Glove, Dependency parsing, LDA, topic modeling, NER, POS tagging, Sentiment analysis etc.
- Experience in AWS EC2 and Google cloud based services.
- Used multiple data visualization libraries and frameworks of python like: Seaborn, Matplotlib plotly, Dash framework and bokeh.
- Experience of using statistical and machine learning algorithms including Regression, instance-based learning, decision trees, Bayesian statistics, Clustering, Ensemble methods.
- Artificial Intelligence ANN, RNN and CNN network using Tensorflow and Keras.
- Frequently using SQL complex queries and Joins for data extraction, data cleaning and processing. Used Mongo dB NoSQL database.
- Strong and innovative approach to problem solving and finding solutions.
- Excellent communicator (written and verbal, formal and informal).
- Image processing, Object Detection and segmentation using OpenCv.

Technical Skills:

Programming Languages	Python, R, SQL, PL/SQL, Apache Spark
Database/ Storage	Oracle Database, Mongo dB, SQLite, Flat files ,MS Excel
Application Tools	Spyder IDE, IntelliJ, Weka Business Intelligence tool, GitHub JIRA, TOAD, Oracle Reports 6i, MS Visio

Industrial Experience:

Larsen and Toubro InfoTech.

(Mar 2018- Present)

Position: Senior Software Engineer

Project Name	Project Details
Semantic Similarity Model on financial data	<p>Built a Semantic similarity model, which is trained on financial data. This model can help the portfolio manager and investors to take the investment decision. Model takes the key phrases and results in the best statements describing the phrase.</p> <p>Algorithm Used: Word2vec, Doc2vec, Stanford Dependency parsing.</p> <p><u>Role and Responsibility:</u></p> <ul style="list-style-type: none">▪ Transfer learning techniques using language modeling ELMO, BERT, and Tensorflow universal encoder.▪ Training word2vec and doc2vec model using data from 300 different companies.▪ Generate the transitional matrix by finding the co-sine similarity score between words of given phrases and sentences.▪ Boosting this score using most similar words and get the final similarity score.▪ Visualize the results using Dash Framework.
Image Processing for Dental Claims	<p>Image processing and classification on human tooth datasets. This model is used for the Insurance company to identify the Insurance claim amount.</p> <p>Algorithm Used: CNN, Keras, OpenCV and Naïve Bayes.</p> <p><u>Role and Responsibility:</u></p> <ul style="list-style-type: none">▪ Perform the labelling on Images using python labellmg library.▪ Created the image classification model on 32 teeth, this model is generating the report whether surgery is required for a given teeth.▪ Model is giving better results compared to the state of the art models.
Aspect and Opinion Mining	<p>Client Name: Principal Financial Group (USA) (Team Size -2)</p> <p><u>Project Description:</u> Created a model for finding the sentiment of customer reviews in terms of aspect and opinion.</p> <p><u>Approach:</u></p> <ul style="list-style-type: none">▪ Unsupervised approach of Sentiment analysis and opinion mining.▪ Comparative study on various sentiment analysis approaches.▪ Created a model using combination of CNN, RNN & Bi directional LSTM.▪ Use data to create BIO tagging using conditional random fields.

10-K Topic Modeling	<p>10-K is the yearly financial report submitted by organizations in US to Security exchange commission. This report contain the detailed about the organization and its management point of view. We used this text information to build a topic model.</p> <p><u>Approach:</u></p> <ul style="list-style-type: none"> ▪ Web-Scrape the data from US SEC Edgar website. ▪ Perform data cleaning, processing and analysis using python libraries. ▪ Apply LDA (Latent Dirichlet allocation) and LDA2Vec algorithm using python genism module. ▪ Generate the results and draw the visualization using pyLDAvis module.
----------------------------	--

TATA Consultancy Services Ltd.

(Jan 2016- Mar-2018)

Position: IT Analyst

Project Name	Project Details
Dis-REC	<p>Regional electricity consumption analysis on consumer's data. Obtain association rules for electricity consumption patterns of regions. Performing area profile for planning residential area aiming electricity efficient living. Prediction of electricity usage in different regions. Algorithm used: Association , Clustering</p> <p><u>Role and Responsibility:</u></p> <ul style="list-style-type: none"> ▪ Use statistical methods to analyze data and generate useful business reports ▪ Use data to create models that depict trends in the customer base and the consumer population as a whole. ▪ Implement the data exploration and data cleaning techniques. ▪ Implement the ML models and show the classification report.
BITS Dissertation	<p>Predictive analytics on Human resource dataset built churn modeling system.</p> <p>Algorithm used like Support vector machine, Naïve Bayes, Decision tree, Random Forest classifier Ensemble learning techniques Bagging Boosting and Stacking.</p> <p><u>Role and Responsibility:</u></p> <ul style="list-style-type: none"> ▪ Perform exploratory data analysis using python libraries. ▪ Recognize the ML model for given dataset. Implement the data mining approach. ▪ Build an Artificial Neural network using Tensorflow and Keras to improve the model performance. ▪ Generate the results and identify the best fit model. ▪ Report the results and present it to higher authorities.

UPS AI	<p>Client Name: United Parcel Service (USA) (Team Size -5)</p> <p>Project Description: UPS, is the world's largest package delivery company and global provider of supply chain management solutions.</p> <p><u>Role and Responsibility:</u></p> <ul style="list-style-type: none"> Data Analyst and developer for UPS Artificial Intelligence team. Developed chat bots for UPS customer's service. <p>Technology: Python, Oracle 11g, Machine learning, AI, Spark.</p>
UPS Post Sales	<p>Client Name: United Parcel Service (USA) (Team Size -10)</p> <p>Project Description: - UPS, is the world's largest package delivery company and global provider of supply chain management solutions. Near real time tracking the vehicle.</p> <p><u>Role and Responsibility:</u></p> <ul style="list-style-type: none"> PL/SQL developer using Oracle 11g. Worked on complex data loading (implemented the batch data cleansing and data loading). Extensively used joins and sub queries to simplify complex queries involving multiple tables. <p>Technology: Oracle 11g, 10g, Unix shell scripting.</p>
The Linde Group	<p>Client Name: The Linde Group (Germany) (Team Size -8)</p> <p><u>Role and Responsibility:</u></p> <ul style="list-style-type: none"> Re-Engineering and Reverse Engineering, Requirement Analysis, Data Analysis, Customization and Maintenance. CRON scripts PL/SQL, SQL programs. BOXI reports and load data to flat files. Developed Functions, Procedures, database triggers, cursors, collections and packages in PL/SQL used to implement data loading business logic. Used DBMS packages like UTL_FILE, DBMS_MAIL to develop PL/SQL programs. ORACLE facilities, SQL*Loader and SQL*PLUS. <p>Technology: Oracle 10g SQL, 11g PL/SQL</p>

Education:

Year	Degree	Institute	Board/Uni.	Result (%)
2017	M.Tech Data Analytics	Birla Institute of Technology & Science Pilani	BITS Pilani	8.82(CGPA)
2012	B.Eng. (Computer Science)	Gyan Ganga Institute of Technology & Sciences, Jabalpur(M.P)	R.G.P.V	77.3%
2008	H.S.C.	Kendriya Vidyalaya G.C.F No1. Jabalpur (M.P)	C.B.S.E	71.0%
2006	S.S.C.	Kendriya Vidyalaya Narsinghpur (M.P)	C.B.S.E	77.6 %