

Ankit Jain

Contact Information

✉ ankit025jain@gmail.com

☎ +91 7676 242 242

<https://www.linkedin.com/in/ankit-jain-07036b71>

<https://github.com/ankit025jain>

Education

Bachelor of Engineering in Computer Science

Dayananda Sagar Academy of
Technology & Management, Bangalore
2012 – 2017

Technical skills

Supervised Machine Learning models

- Linear Regression
- Logistic Regression
- Naïve Bayes
- Decision Trees
- Random Forest
- Support Vector Machine
- Xgboost and Adaboost
- Neural Networks
- Convolution Neural Networks
- Recurrent Neural Networks
- LSTM
- Text Mining

Unsupervised Machine Learning models

- Dimensionality Reduction Techniques (PCA & SVD)
- Clustering
- Association Rule Mining
- Embeddings (Glove, word2vec, ELMO)

Programming Languages

- Python
- R
- C

Tools

- Tableau
- Prodigy
- PowerBI
- Monarch Datawatch
- Microsoft Excel

Professional Summary

- Worked on a variety of projects in Natural Language Processing (NLP). Diverse knowledge in NLP (Text domain).
- Very good conceptual and practical understating of machine learning algorithms, Deep learning algorithms and Big data concepts.
- Excellent knowledge and understanding on how to scrape the data from various websites using BeautifulSoup library.
- Excellent knowledge in Visualization using Tableau, python (plotly, seaborn, matplotlib).
- Worked on a project – **Smart highway using Arduino with Data retrieval on iOS app.**

Certifications

- Post Graduate Program in **Big Data Analytics & Optimization** from **International School of Engineering**
The program is certified for the quality, pedagogy, and assessment by LTI of Carnegie Mellon University, USA.

Data Science Projects at INSOFE

Aspect based sentiment analysis for a leading automobile car industry client in US.

To predict if the different aspects of the car have a positive response or negative response based on the reviews on twitter, YouTube and other social media platforms and verbatims file shared by the client. Apart from web scraping the reviews the main task was to provide overall sentiment and provide sentiment based on the different aspects defined by the client.

- Techniques: **Text preprocessing, CNN, LSTM, (various deep learning architectures), word2vec, Glove, ELMo embedding, spacy, spacy(Dependency parsing).**
- Languages: Python
- Summary: Web scraped the data from Twitter, YouTube using beautifulsoup, Cleaned the text using text preprocessing techniques (Tokenization, Stemming, Lemmatization, Regex). Read various papers on aspect based sentiment analysis (ABSA) and came up with some amazing rules that helped in capturing the aspects and their sentiments from the extracted text using Spacy dependency parsing.

Automation of Resume parsing for a US client.

To automate the process of resume parsing based on the profile and job requirements specified by the client. The target was to predict if the resume of the person is eligible based on different criteria specified by the client.

- Techniques: **Spacy, prodigy, Regular Expression, Models – Logistic Regression, Naïve Bayes, Decision Trees, Random Forest, PCA, SVM, XGBoost, MLP, CNN, RNN, LSTM, Glove embeddings.**
- Languages: Python
- Summary: Read the resume files, cleaned the text, tagged the resume with different labels using prodigy tool and also used regular expression to separate different sections of the resume. Automated the parsing of resumes of candidates based on the different criteria specified by the client, applied various models and were able to automate the process with 82% accuracy.

Databases

- SQL

Deployment

- Docker

Big Data Technologies

- Hadoop Ecosystem
- Spark

Interests

- Yoga
- Gym

Aspect extraction, sentiment prediction and deployment of hotel recommendation engine for a Hotel industry client.

Web scrape the reviews from online sources (tripadvisor.com, booking.com). Analysis of the reviews and applied various techniques to identify the different aspects of the reviews and sentiment analysis of the reviews and recommendation engine using deep learning for hotel recommendation.

- Techniques: **Beautiful soup, Selenium, Deep learning, Spacy, Spacy(dependency parsing).**
- Languages: Python
- Summary: Extracted the data from various websites using beautiful soup and used spacy's dependency parsing to identify the different aspects and sentiment for these aspects using Vader sentiment analysis and textblob sentiment libraries. Combined all the reviews for a hotel and then summarized the hotel based on the reviews using text summarization, build and deployed the recommendation engine on docker.

(Text Mining) Consumer compliance

To cluster consumer grievances as "Un-fair and deceptive business practice related (UFDP)" and "efficiency related (not UFDP)". And predict the ratings for banks based on these grievance responses

- Techniques: **Logistic Regression, Naïve Bayes, Clustering, Decision Trees, Random Forest, PCA, SVM, XGBoost, MLP, LSTM, Glove embeddings.**
- Languages: Python
- Summary: Performed various text preprocessing and applied various models and found **random forest** to be an appropriate model for this business case producing the highest accuracy.

Workshop Attended

- Attended **workshop on IOT, Ethical Hacking**
- Attended **workshop on Strategic Management, Leadership and Behavioral Skills**, conducted by **Centre for innovation and learning.**

Experience – Overall 2 years

INSOFE

Designation : Data Scientist

Duration : December 2018 – Present

Roles and Responsibilities

- Worked on consulting projects for a leading US client in text domain for **sentiment analysis and aspect based sentiment analysis, Recommendation Engine, Text summarization.**
- Worked on Automation of Resume Parsing for a leading recruitment company in US. (Tools used: Spacy, Prodigy, Text mining tools, Regular expression).
- Web scraped the data from various websites and performed ABSA, text summarization and build recommendation engines.
- Working on a research project (An Unsupervised Neural Attention Model for Aspect Extraction).
- Helped many corporate clients setup and train their data science teams.
- Helped Data Scientists at Insofe in delivering contents on ML, DL, Visualization and Big data for lab sessions.

Ernst and Young LLP

Designation : Risk Analyst, Advisory services.

Duration : July 2017 – December 2017

Roles and Responsibilities

- Driving Engagements throughout IT Audit lifecycle*
- Initial planning, walkthrough, ITGC testing for Manage Access, Manage Change and Manage IT Operations for test of operating effectiveness.