



Praveen Bushipaka

● ABOUT ME

Researcher in ML/AI with experience in working with Large Language Models (LLM), with interests in multi-disciplinary fields. Additionally understanding in statistics, analytical techniques. Highly organized, motivated and diligent with significant background in Research in AI and Data science.

● WORK EXPERIENCE

01/09/2022 – CURRENT Pisa, Italy

RESEARCH ASSOCIATE SCUOLA SUPERIORE SANT'ANNA

- Research on AI in Italian Law to fasten the process at judicial offices.
- Collaborated with a five-person team to develop a novel Extractive Summarization method using Italian-LEGAL-BERT (LLM) to automate summarization of legal documents.
- Role classification of sentences with a LLM as encoder on a Hierarchical BERT architecture.
- Current research on developing a LLM for legal domain with a LoRa for efficient training and less computational power.

20/09/2021 – 01/01/2022 Rome, Italy

DATA SCIENTISTS BHAVE S.R.L

- Behavior data analysis of customers and Brand awareness for clients product, and identifying KPIs which drives traffic towards clients products using ML algorithms
- Brand comparison with competitors and evaluated clients market share
- To evaluate customer satisfaction, developed net promoter score (NPS) on scale of 10
- Development of Machine Learning model with Random Forest to predict suitable therapy to patients
- Used SHAPely algorithm to find features influencing model and presented them to stakeholders about features
- Read research papers to find optimal approach in developing Machine Learning Model

02/12/2019 – 12/02/2021 Milan, Italy

DATA SCIENTISTS WUNDERMAN THOMPSON ITALY

- Customer Segmentation on 78,000 customers with clustering to implement personalized marketing strategy.
- Worked on complex data sets consisting of millions of rows of customer transactions data and variables more than 100 Data analysis and reporting to improve customer experience and customer support system.
- Data collection with MySQL and Extracted insights, KPI's from big data to make data driven decisions.
- Developed Market Basket Analysis model using Apriori Algorithm to find association among products and optimize product placement.
- Compiled, cleaned and manipulated data for proper handling
- Developed polished visualizations in Tableau to share results of data analyses

● EDUCATION AND TRAINING

02/09/2017 – 16/10/2020 Italy

MASTERS IN SCIENCE IN COMPUTER SCIENCE University of Milan

- Topics included in the course
- - Deep Learning (CNN, RNN, LSTM, GRU)
- - Machine Learning (Python, R)
- - Information Management (Data mining, statistics, clustering and classification)
- - Machine Vision (Image recognition in industry with silicon software using CNN)
- - Information Retrieval (NLP techniques, traditional ML and transformers)
- - Bioinformatics
- - Digital Image processing
- - Wireless Networks
- - Business Intelligence (BPMN, Value Model)
- - Project Management (UML, WBS)

Website <https://www.unimi.it/en/education/computer-science-master> | **Final grade** 98/110 | **Type of credits** ECTS |

Number of credits 120 | **Thesis** INSIGHT EXTRACTION FROM DATA TO IMPROVE CUSTOMER EXPERIENCE

23/09/2013 – 01/05/2017 Hyderabad, India

BACHELORS IN TECHNOLOGY IN INFORMATION TECHNOLOGY Sreenidhi Institute of Science and Technology (affiliated by Jawaharlal Nehru Technical University)

Website <https://www.sreenidhi.edu.in/pagecontents.php?catid=4&scatid=408>

● LANGUAGE SKILLS

Mother tongue(s): **TELUGU**

Other language(s):

	UNDERSTANDING		SPEAKING		WRITING
	Listening	Reading	Spoken production	Spoken interaction	
ENGLISH	C2	C2	C2	C2	C1
ITALIAN	A2	B1	B1	B1	A2

Levels: A1 and A2: Basic user; B1 and B2: Independent user; C1 and C2: Proficient user

● DIGITAL SKILLS

Programming languages (Python and R program) | Relational Databases: SQL | Flask (beginner) | Data extraction, Data preprocessing, Data visualization, Data cleaning | Machine Learning | Deep Learning | Pandas, Numpy, Scikit-learn, Scipy libraries | Team-work oriented | Analytical skills | Machine Learning, NLP | Multilingual NLP | Deep Learning frameworks: PyTorch, Tensorflow, Keras

● ADDITIONAL INFORMATION

SOCIAL AND POLITICAL ACTIVITIES

Hyderabad, India

Statistical analysis page on Instagram

- Started a Instagram Page which depicts statistics of Dalits who are marginalized community in India
- Data collected from Indian Government data portal
- Developed data visualization to demonstrate the results.

PROJECTS

End to End Data Science Project | Python

- **Problem Statement:** Predict the Web article genre (news, entertainment, publication etc) based on the url and content
- Data is collected from 53000 web articles
- Each article is categorized into their respective genres (news, publication) in a total of 9 genres
- TFIDF to vectorize the data
- The problem is a supervised prediction, applied Logistic regression, Decision trees, Random forest and Naive Bayes
- Cross validation to not over-fit data
- Evaluation metric F1-score to find the best performing model
- Pickled the best performing model and developed the backend API with Flask for the deployment
- Deployed on heroku
- <https://web-page-classification.herokuapp.com/>

ORGANISATIONAL SKILLS

Organisational Skills

- Researching, Analytical thinking
- Ability to work in teams in multi-cultural environments and in teams.
- Good ability in planning and socialization skills coming from experiences in working.

COMMUNICATION AND INTERPERSONAL SKILLS

Excellent communication skills in a multicultural workplace

- As a researcher in inter-disciplinary field, I collaborated with different departments at Sant'Anna.
- My experience at Wunderman Thompson Italy, helped me improve my communication skills in an international working environment.
- Apart, my fellow peers who are International students also gave me confidence, diplomacy when working in international environment.