Dhaval Taunk

Portfolio: *dhavaltaunk08.github.io*
Github: *github.com/DhavalTaunk08*

Email: taunkdhaval08@gmail.com
Mobile: +91-9713899088
LinkedIn: *linkedin.com/in/dhavaltaunk08/*

Education

- **International Institute of Information Technology** — Hyderabad, Telangana, India
  MS by Research - Computer Science and Engineering; CGPA: 10/10 — *Aug 2021 - Present*

- **Indian Institute of Information Technology, Design and Manufacturing** — Jabalpur, M.P., India
  Bachelors of Technology - Computer Science and Engineering; CGPA: 7.2/10 — *Aug 2016 - June 2020*

Experience

- **International Institute of Information Technology, Hyderabad**
  Research Assistant - Information Retrieval & Extraction Lab (iREL) — *Jan 2022 - Present*
  - **Role**: Pursuing MS by Research under the guidance of Prof. Vasudeva Varma and Prof. Manish Gupta.
  - **Work**: Working as Research Assistant on different problems like text generation for low resource languages, outline generation for encyclopedic text, question answering using commonsense reasoning.

- **Yes Bank**
  Data Scientist — *Aug 2020 - July 2021*
  - **Industry and Sub-industry Classification**: The project aims to identify potential small scale industries as customers and help them in growing their business.
  - **Loyalty Rewards Program**: Worked on Loyalty Rewards program which aimed at awarding reward points to customers based on their transactions and given set of rules.

- **Jio Haptik Technologies Limited**
  Machine Learning Intern — *May 2019 - Nov 2019*
  - **Project**: Build an intent detection system of chatbots by finetuning and testing several deep learning based models s like BiMPM, ABCNN, BERT, Siamese based networks, USE, ULMfit, tf-idf etc.
  - **Outcome**: Improved bot's performance by using ULMfit and tf-idf and thereby leading to a 13% (approx.) rise in accuracy and an enhanced customer experience.

- **IIT Guwahati**
  Summer Research Intern — *May 2018 - July 2018*
  - **Project**: Implemented Gender Classification by using deep neural networks in live video streaming by training 3 different models on image, audio and video dataset.
  - **Outcome**: The achieved accuracy of models is 87.4%, 98.7% and 68.4% for the image, audio files and video files respectively.

Publications

1. Dhaval Taunk, Shivprasad Sagare, Anupam Patil, Shivansh Subramanian, Manish Gupta, and Vasudeva Varma. *XWikiGen: Cross-lingual Summarization for Encyclopedic Text Generation in Low Resource Languages.* In *Proceedings of the Web Conference 2023*, WWW '23, New York, NY, USA, 2023. Association for Computing Machinery

2. Dhaval Taunk, Lakshya Khanna, Pavan Kandru, Vasudeva Varma, Charu Sharma, and Makarand Tapaswi. *GrapeQA: GRaph Augmentation and Pruning to Enhance Question-Answering.* In *Companion Proceedings of the Web Conference 2023*, NLP4KGC (WWW '23), New York, NY, USA, 2023. Association for Computing Machinery

3. Dhaval Taunk and Vasudeva Varma. *Summarizing Indian Languages using Multilingual Transformers based Models.* In *Forum for Information Retrieval Evaluation, December 9-13, 2022, India*, 2022

4. Sagar Joshi, Dhaval Taunk, and Vasudeva Varma. *IIIT-MLNS at SemEval-2022 Task 8: Siamese Architecture for Modeling Multilingual News Similarity.* In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, pages 1145–1150, 2022

5. Dhaval Taunk, Sagar Joshi, and Vasudeva Varma. *Profiling irony and stereotype spreaders on Twitter based on term frequency in tweets.* In *Conference and Labs of the Evaluation Forum (CLEF) 2022*, 2022

Skills Summary

- **Skills**: Deep Learning, Machine Learning, Natural Language Processing, Algorithms
- **Tools & Frameworks**: Python, PyTorch, Keras, Scikit-Learn, MySQL

Projects

- **Wikipedia Search Engine**: The aim of this project is to build a search engine from scratch. For this, English Wikipedia dump of size 84GB was used to create index. Then, search functionality was implemented using TF-IDF based ranking mechanism. Tech: Python, XML, NLTK, PyStemmer.
- **Question Answering using CommonSense Reasoning**: Improved performance by adding modifications (PEGA and CANP) on the model proposed by QAGNN on datasets CSQA, OBQA, MedQA which involves training language models and graph neural networks simultaneously.
- **Text Segmentation in images using Auto-encoders**: This projext aimed to create a system that can perform text segmentation in images using Auto-encoders. For training purpose, KAIST Text Scene dateset was used. Tech: Python, Keras, OpenCV.
- **Salient Object Detection**: The objective of this project is to perform Salient Object Detection by implementing a paper called *Deep Embedding Features for Salient Object Detection*.

Honors and Awards

- Achieved 2^{nd} position in the shared task Indian Language Summarization organized in FIRE 2022.
- Secured a rank of 288 out of 6871 candidates in Capgemini Tech Challenge (Data Science) 2018.
- Awarded Meritorious student incentive on scoring above 85% in class 12^{th} by Madhya Pradesh Govt.

Volunteer Experience

- **Teaching Assistant for IRE course**: Worked as a teaching assistant for the course Information Retrieval & Extraction (IRE) in Monsoon 2022 semester @ IIIT Hyderabad and mentored 8 students in their course project.
- **Writer @ AnalyticsVidhya/Medium**: Wrote technical articles related to machine learning, deep learning etc. fields for Analytics Vidhya publication on *Medium*
- **Contributor @ HuggingFace's Transformers**: Contributed 2 Community notebooks in HuggingFace's *Transformers* repository related to multi-class and multi-label text classification