Gunjan Chhablani

chhablani.gunjan@gmail.com gchhablani gchhablani gchhablani.github.io +91-7719900640

Employment

Oracle India Pvt. Ltd. **Feb 2020 – Present**
Applications Engineer *Bangalore, KA, India*

- Oracle Cloud AI Services - Language
- Actively involved in development of the text classification service.
- Maintaining the training and serving pipelines for custom text classification and intent detection models.
- Python, Docker, TensorFlow, git.

Oracle India Pvt. Ltd. **Sep 2020 – Feb 2020**
Applications Engineer *Hyderabad, TG, India*

- Oracle Fusion Applications - HCM Cloud development.
- Actively involved in UI and Model(View/Entity) layer implementation - development and unit testing - of various modules in Global HR - Oracle HCM Cloud.
- Identifying process automation opportunities and automating various tasks including REST API testing.
- Worked as Sprint Master for Project Progress and Release Management. Managed Product and Sprint Backlog to achieve the Sprint objectives in agile methodology.
- Oracle ADF, Java, SQL, PL/SQL, ADE, Shell (csh, bash), RESTful Services, JDeveloper 11g/12c, Python, JavaScript.

Internships

University of Pittsburgh **Jan 2020 – Aug 2020**
Student Intern *Pittsburgh, PA, USA*

- Studied the performance of novel weight initialization technique in neural networks - Perceptually Inspired Deep Learning (PIDL) on MNIST/FashionMNIST and OCT segmentation datasets.
- Used U-Net/U-Net++ deep learning architectures with VGG-19 backbone on a home-grown Choroid segmentation dataset in OCT scans.
- Abstract accepted at ARVO, 2021. Another paper in the pipeline for publication at TVST, ARVO.

IBM India Pvt. Ltd. **May 2019 – Jul 2019**
Project Trainee *Gurgaon, HN, India*

- Actively contributed to the ongoing project at IBM Watson Services on Instance Segmentation.
- Annotated car parts and damages on 100+ insurance claims using VGG Image Annotator.
- Helped in developing a MaskRCNN model in TensorFlow for car parts segmentation and damage detection.
- Achieved 85% Mean Average Precision score on car parts segmentation.

L.V. Prasad Eye Institute - Winter Intern **Dec 2018 – Jan 2019**
Winter Intern *Hyderabad, TG, India*

- Segmented Hard-Exudates from Real-life OCT Scans on home-grown dataset with 100 images.
- Segmented choroids in 900+ OCT scans using Tensor Voting based segmentation tool in MATLAB.
- Built networks like U-Net/Deep U-Net in Keras for segmentation of hard-exudates with Dice Coefficient as evaluation metric.

Education

BITS Pilani, K.K. Birla Goa Campus **Aug 2016 – Sep 2020**
*B.E. Computer Science - **CGPA** - **9.65/10** - Batch Rank 3/617* *Goa, India*

Publications

Superpixel-based Knowledge Infusion in Deep Neural Networks for Image Classification **Paper Link**
Accepted at ACMSE 2022

- Proposed a technique involving fusing knowledge from superpixel-based graphs with convolutional neural networks for image classification tasks.
- Analyzed the performance of tate-of-the-art approaches like Graph Neural Networks, Graph Attention Networks with pre-trained Convolutional Neural Networks on several image classification datasets.

DRIFT: A Toolkit for Diachronic Analysis of Scientific Literature **Paper Link**
Published at EMNLP 2021 (Demo Track)

- Actively developed an open-source tool with Streamlit-backed GUI for diachronic analysis of scientific literature.
- Compiled and implemented 10 different kinds of analysis techniques including semantic drift, acceleration analysis, wordclouds, trend detection, etc.
- Performed extensive analysis on arxiv.cs.CL dataset. Paper accepted at EMNLP Demo Track 2021.

Multitask Prompted Training Enables Zero-Shot Task Generalization **Paper Link**
Accepted at ICLR 2022 (Spotlight)

- Proposed a prompting approach towards zero-shot generalization in NLP domain using existing datasets in the HuggingFace hub.
- Added prompt templates for many datasets using Streamlit and Jinja templating as a part of the BigScience open-source workshop.

Datasets: A Community Library for Natural Language Processing **Paper Link**
Published at EMNLP 2021 (Demo Track) - Best Paper

- The demonstration depicts a powerful library - HuggingFace datasets where several NLP datasets can be accessed in various settings with simple APIs.
- Actively contributed over 11 datasets to the HuggingFace datasets.
- Added utilities for dataset metadata validation and tests to the repository.

SemEval-2021 Task 5: Toxic Spans Detection **Paper Link**
Published at SemEval Workshop, ACL-IJCNLP 2021

- Worked on toxic spans detection where the offsets of the toxic phrases are to be detected in a given text.
- Performed a thorough analysis of token classification, span prediction, and novel hybrid approaches involving LSTM, CRF and BERT.
- Achieved an offset-wise F1 score of 67.53% (rank 28/92) officially, and 68.95% during the post-evaluation period.

SemEval-2021 Task 4: Reading Comprehension of Abstract Meaning **Paper Link**
Published at SemEval Workshop, ACL-IJCNLP 2021

- Worked on a cloze-style question answering task with abstract words as options.
- Analyzed the working of existing state-of-the-art techniques like BERTCloze, GAReader, and related variants in PyTorch.
- Improved the performance using engineering features for imperceptibility and using augmentation and pretraining and hyponyms for non-specificity.
- Achieved a score of 77.28% (rank 15/23) on imperceptibility, 78.98% (rank 17/28) on non-specificity subtasks. Paper accepted at SemEval2021 workshop (ACL-IJCNLP 2021).

Open Source Contributions

HuggingFace **Feb 2020 – Present**
Core Contributor

- Active core contributor to HuggingFace transformers and datasets libraries.
- Contributed and maintaining the VisualBERT, PLBart and FNet models in PyTorch. Currently working on adding SMITH and OFA models to the repository.
- Contributed over 11 datasets and worked on metadata validation for the datasets library. Currently working on automated metadata generation for datasets. Paper accepted at EMNLP Demo Track 2021.

Personal Projects

Multimodal Multilingual Models | JAX/Flax, PyTorch, Python, TPUv3-8 VM **Jul 2021 – Aug 2021**

- Participated with one other team member in the HuggingFace JAX/Flax sprint for multilingual image captioning and multilingual visual question answering (VQA).
- Created a novel combination of ViT encoder-BART decoder for image captioning and ViT encoder+BERT encoder in Flax for VQA.
- Worked on CC12M image captioning dataset translation using MarianMT and MBART50 models and trained the models on TPUv3 VMs provided by Google.
- Prepared a Streamlit demo for both shared on HuggingFace spaces. The project was ranked among top-15 out of 50 other teams.

Detecting Terrorism Promotion on Twitter | Python, Jupyter Notebook, SQL **Aug 2019 – Aug 2020**

- Extracted context-based and socio-psychological features from the data using LiWC, GrammarBot, and TextRazor API after pre-processing and training models (XGBoost, Neural Networks, Linear Regression) on the data.
- Used semi-supervised learning to make a mental-state model to identify promotion of terrorist sympathizers, and a categorization method to classify users based on their level of involvement.

Technical Skills

Languages: Python, Java, C, C++, SQL, Shell
Technologies/Frameworks: Git, PyTorch, Keras, Tensorflow, JAX/Flax, ADF, Scikit-learn, Streamlit

Additional Info

Scholarships: *Institute Merit Scholarship* - for ranking among top 1% of the batch for 8 semesters at BITS Goa, *National Talent Search Scholarship* 2014 – by NCERT for qualifying National Talent Search Examination.
Certificates: Deep Learning Specialization (Coursera), Machine Learning (Coursera), Intro to Data Science in Python (Coursera), Applied Machine Learning in Python (Coursera), Algorithmic Toolbox (Coursera).
Teaching Experience:
Instructor of ML/DL Course Course : Intro to ML and DL under the Center for Technical Education at BITS, Goa.
Teaching Assistant : Theory of Computation — Data Structures and Algorithms — Discrete Structures for Computer Science.