

# Harshit Pandey

Portfolio: [harsh4799.github.io](https://harsh4799.github.io)

Github: [github.com/harsh4799](https://github.com/harsh4799)

Email: [hp2pandey1@gmail.com](mailto:hp2pandey1@gmail.com)

Mobile: +91 96893-41271

Google Scholar: [scholar.google.com/harshit](https://scholar.google.com/harshit)

## EDUCATION

---

- **Savitribai Phule Pune University** Pune, India  
*Bachelor of Engineering - Computer Engineering; CGPA: 9.08/10* *July 2017 - July 2021*  
*Relevant Courses: Object Oriented Programming, Data Structures, Analysis Of Algorithms, Databases, Computer Architecture, Theory of Computation, Networking, Operating Systems, Machine Learning, Artificial Intelligence, Data Analytics, Data Mining & Warehousing*

## SKILLS SUMMARY

---

- **Languages:** Python, Java, JavaScript, C++, SQL, Bash
- **Frameworks:** PyTorch, NodeJs, Springboot, ReactJs, Angular
- **Tools:** Git, PostgreSQL, MySQL, Kubernetes, Docker
- **Platforms:** Linux, AWS, Web, Windows

## WORK EXPERIENCE

---

- **Cognizant Technology Solutions** Full-Time  
*CDB-FSE Programmer Analyst* *Aug 2021 - Current*
- **Cognizant Technology Solutions** *Mar 2021 - July-2021*  
*FSE Intern*
  - **Audit Management System:**
    - Created 4 microservices with Springboot that included security microservices and feign client microservices.
    - Set up post-commit CI/CD pipelines that automatically tested and deployed the clusters to AWS ECS.
- **Merkle Sokrati** *Feb 2020 - Aug 2020*  
*Software Design Engineer Intern*
  - **Interactive Voice Response (IVR) Management Console:**
    - Added Redis caching along with security mechanisms with JWT tokens and reCAPTCHA.
    - Highly customizable React.js based console with options for parallel calling and schedulable numbers.
    - Backend technology used was Node.js and PostgreSQL.
  - **Vtiger Migration:**
    - Migrated data from the Manyavar's old CRM using AWS Lambda scripts in conjunction with Vtiger's API.
    - Created a script that puts new leads into AWS's SQS and migrates it with AWS Lambda.
  - **Invoice Automation:**
    - Created web-based portal that would automate processing of Invoices.
    - Created with React.Js and Node.Js

## PUBLICATIONS

---

- **DRIFT: A Toolkit for Diachronic Analysis of Scientific Literature** *Accepted: Aug 2021*  
*Empirical Methods in Natural Language Processing (EMNLP-21)*
  - Analysing change in semantic meaning of words over time using diachronic word embeddings.
  - Collated a variety of analysis methods to monitor trends and patterns of development in scientific literature.
  - Created a user-friendly customizable web-demonstration that provides an extremely easy-to-use interface for studying trends and patterns of development in scientific literature.
  - Supports customizable training of diachronic word embeddings with TWEC model.

[\[Paper\]](#) [\[Code\]](#) [\[Demo\]](#)
- **Improving Abstract Reading Comprehension using Augmentation, Ling. Features & Voting** *Aug 2021*  
*International Workshop on Semantic Evaluation (SemEval@ACL-IJCNLP)*
  - Given a fill-in-the-blank-type question and a corresponding context, the task is to predict the most suitable word from a list of 5 abstract options.
  - Experimented with many pretrained MLM models, like BERT, ALBERT, DistilBERT and RoBERTa.
  - Using WordNet, added linguistic features to improve task performance.
  - Used integrated gradients to explain predictions on samples.

[\[Paper\]](#) [\[Dataset\]](#) [\[Code\]](#)

- **Toxic-Spans Detection with BERT-based Token Classification & Span Prediction Techniques**  
*International Workshop on Semantic Evaluation (SemEval@ACL-IJCNLP)* *Aug 2021*
  - Given a passage, the task is to mark the spans of text that are toxic.
  - Performed thorough ablative analysis testing existing methods and creating novel techniques.
  - Used variations of BERT based models with and without Conditional Random Field(CRFs), multi-spans classification and offset prediction method.
  - Used integrated gradients to explain predictions on samples.

[\[Paper\]](#) [\[Dataset\]](#) [\[Code\]](#)

- **Document Ranking with XLNet-Based Models**  
*Text REtrieval Conference (TREC-2020)* *Nov 2020*
  - Given transcripts of 100,000 podcasts (15+ GB of Data) the task was to predict the top 1000 relevant results from a user's query.
  - Surpassed the baseline using BM25, a traditional IR technique with XLNet, a transformer based model.
  - Devised a contextual representation approach of the former model for faster inference times.

[\[Paper\]](#) [\[Dataset\]](#)

- **Superpixel-based Domain-Knowledge Infusion in Computer Vision**  
*Pre-Print* *May 2021*
  - Hypothesised that information carried out by graph representations of superpixels could improve results of traditional vision models.
  - Experimented with hybrid models that leverages CNNs and Graph Convolutional Networks, to deal with relational information. superpixel information
  - Observed an increment in accuracy in most datasets in consideration.

[\[Paper\]](#) [\[Code\]](#)

## PROJECTS

---

- **Adversarial Deep Learning Book:** Research oriented, open source, tutorials for upcoming book on adversarial deep learning. Project is part of SAIDL I am one of the lead developers of the team. (Current) [\[Code\]](#)
- **Multitask Prompt Tuning Enables Zero-Shot Task Generalization:** A project from Hugging Face's Big Science Workshop that involves the study of improving zero-shot performance via the means of prompt tuning. (Current) [\[Code\]](#)
- **Alternative Crop Recommendation System:** Created an end-to-end website application (MERN Stack) for farmers that recommends alternative crops with machine learning models, given multiple factors based on location and soil conditions. Project created for IEEE National Hackathon. (Dec '20) [\[Code\]](#)
- **DecepticonNLP:** Open source library for robustness monitoring & adversarial debugging of NLP models.(Jul '20) [\[Code\]](#)

## CERTIFICATIONS

---

- **AWS Certified Cloud Practitioner** AWS  
*Can effectively demonstrate an overall knowledge of the AWS Cloud independent of a specific job role.* *Sep 2021*
- **The Complete Web Developer in 2020: Zero to Mastery** Udemy  
*Comprehensive course on web development with technologies like React.js and Node.js.* *Feb 2020*
- **Deep Learning Specialization** Coursera  
*Break into the field of AI, with comprehensive courses on Neural Network, Computer Vision and NLP.* *Aug 2019*
- **Machine Learning QSTP Course** BITS  
*Introduction to machine learning with a final project.* *Jun 2018*

## VOLUNTEER EXPERIENCE

---

- **Core Member of Language Research Group** BITS  
*Weekly paper reading sessions, research discussions and opportunities for research projects.* *Apr 2020 - Present*

## ACHIEVEMENTS

---

- **IEEE National Gov-TechThon:** One of the top 20 teams from task-1
- **Smart India Hackathon:** Team lead of one of the top 6 teams from the internal hackathon.
- **Departmental Project Competition:**First place in best departmental project.