

Shashwat Pandey

925-367-1860 | spandey7@ucsc.edu | [linkedin.com/in/shashwat1225](https://www.linkedin.com/in/shashwat1225) | github.com/shashwat1225

EDUCATION

University of California

Master of Science in Natural Language Processing

Santa Cruz, CA

Sep. 2022 – Dec. 2023

- Grade: 4.0+/4.0

- Coursework: Machine Learning, Deep Learning, Conversational Agents, Machine Translation, Generative Transformers, Information Retrieval, Attention Networks, Language Modelling, Statistical Modelling

Jaypee Institute of Information Technology

Bachelor of Technology in Electronics and Communications Engineering

Noida, India

Aug. 2016 – July 2020

- Coursework: Algorithms & Complexity Theory, Artificial Intelligence, Probability, Statistics, Linear Algebra, Computer Architecture

EXPERIENCE

Software Engineer

Amdocs DVCI

Dec. 2020 – Aug. 2022

Pune, India

- Delivered data science solutions for AT&T, automating and optimizing daily operations across various domains, resulting in a 30% increase in efficiency and a 20% reduction in manual tasks for relevant business units.
- Pioneered a churn prediction model for predicting the customer churn using Decision tree with 91% accuracy and assisted the marketing team to build effective customer retention strategies, mitigating churn by 22%
- Conceptualized and created a knowledge graph database of news events for AT&T, extracted from tweets using Java, Virtuoso, Stanford CoreNLP, Apache Jena, RDF.
- Developed a CLV model increasing customer lifetime value by 15% and implemented a Churn Detection model with F1 of 0.85 and AUC-ROC score of 0.92, reducing churn by 25%, and increasing retention from 78% to 84%. This integration improved customer segmentation and campaign ROI by 20%.
- Collaborated with business teams to migrate 50+ billion subscriber records along with 30+ systems data integration on Azure cloud.

Associate Data Scientist

Datakalp LLP

Oct. 2020 – Dec. 2020

Bengaluru, India

- Independently developed a Fire Detection Machine Learning model with 95% recall and 99.7% precision rate
- Demonstrated convoluted statistical solutions to senior business stakeholders and clients of global stature independently

ACADEMIC PROJECTS

Dialogue Dataset for Natural Language Generation Engine

- Developed an instruction-based fine-tuning method to generate a text dataset of 5000 dialogues that is semantically accurate and coherent with 9 dialogue acts(DA).
- Investigated with 10 instruction-based prompts over 3 baseline Large Language Models(LLMs) - **GPT Neo, GPT-3.5, and Jurassic.**
- Generated a dataset with **BLEU score of 0.65** and DA Classifier **recall value of 0.86.**

Interactive Neural Machine Translation Engine

- Pioneered a novel Interactive-Predictive Neural Machine Translation (IPNMT) system that allows real-time interaction between human translators and machine translation systems
- Integrated sampling-based decoding methods for interactivity with given LLMs - **Google's T5 and mT5 from huggingface.**
- Achieved best **BLEU score of 0.58** on mT5 model with **20% partial translations**

Using OCR to automate document conversion to LATEX

- Pioneered a method to modularize and compartmentalize OCR tasks that each can focus on the different types of issues that can occur at different levels of granularity.
- Ran up to 5 feature extraction techniques and 10+ permutations of EDA over 2 machine learning models: **KNN and Logistic Regression**
- Came up with the best **classification scores of 0.951** on logistic regression and **0.946** on KNN under the best hyperparameter configurations

Explainable Detection of Online Sexism

- Performed Explainable Detection of Online Sexism (EDOS) for detecting sexist content with a novel hierarchical taxonomy and 11 different fine-grained labels.
- Investigated 4 different techniques from using elementary machine learning models such as **random forest, XGBoost to Encoder/Decoder based RNNs** and then tweaking them up with **Attention-based LSTMs.**
- Successfully generated **F1 score of 0.465 (among top 20 in the competition).**

PUBLICATIONS

S. Pandey and A. Rohatgi, "Using OCR to automate document conversion to LATEX," 2021 International Conference on Computational Intelligence and Computing Applications (ICCICA), Nagpur, India, 2021, pp. 1-8, doi: 10.1109/ICCICA52458.2021.9697266.

TECHNICAL SKILLS

Languages: Python, C/C++, SQL (Postgres, NoSQL, SPARQL), R

Frameworks: PyTorch, TensorFlow, Keras, HuggingFace Transformers, Flask, MLFlow, REST API, Gradio, Modal, NLTK, spacy

Developer Tools: Git, Docker, Kubernetes, Jenkins, Grafana, Prometheus, Azure, AWS

Competencies: NLP, Computer Vision, Time Series Analysis, Reinforcement Learning, Autoregressive Models, Parallel Programming, MLOps