John Henry Cruz

johnhenrycruz000@gmail.com · 832-782-7523 · https://johnhenrycruz.netlify.app/

EXPERIENCE

AT&T Chief Data Office, FraudML Team July 2022 - Present
Technology Development Program - Engineer I Data Analyst Los Angeles, CA

- Produced a script to collect voice biometric data related to speaking styles to be used as model features
- Developed a fraud offender speaker verification model using NVIDIA NeMo to build a feature used in a fraud prediction model
- Designed the pipeline for transferring audio files and transcripts from call centers for analyses securely to the cloud
- Conducted a topic modeling analysis on call center transcripts to build a model to predict fraud-related activities from callers

The University of Texas at Austin, Pennebaker Language Lab September 2020 – January 2022
Undergraduate Researcher Austin, TX

- Studied group therapy transcripts for relationships between talking speed/hesitation with different psychological dimensions
- Analyzed survey and Reddit data concerning COVID-19 using LIWC, MEM, and other computational linguistics tools
- Predicted and interpret major themes and feelings during the pandemic such as Masks through statistical models and PCAs
- Developed dictionaries on different psychological dimensions and created text datasets for the new version of LIWC.

The University of Texas at Austin, Jessy Li Computational Linguistics Lab August 2021 – December 2021
Undergraduate Researcher Austin, TX

- Built and trained different baseline models to categorize various texts to a distinct sentiment
- Developed Amazon Mechanical Turk tasks using JavaScript and HTML for annotations of emotions in COVID-19 data
- Scraped and Analyzed COVID-19 data from support subreddits and 16 cities subreddits for emotional triggers

Northwestern University, Creative Cognition Lab June 2021 – July 2021
Summer Research Opportunity Program (SROP) Research Intern Evanston, IL

- Developed algorithmic methods to automate the scoring process of AUT Metrics to tackle subjectivity with human scoring
- Coded these methods using Python, semantic spaces, and clustering methods to calculate novelty, originality, and flexibility
- Presented findings and analysis methods at the culminating Northwestern SROP Research Symposium

PUBLICATIONS

Yu, Y., Beaty, R., Forthmann, B., Beeman, M., Cruz, J. H., & Johnson, D. R.. *A mad method to assess idea novelty: Improving validity of automatic scoring using maximum associative distance (MAD)*. https://doi.org/10.31234/osf.io/vgxpk
May 2022

- Designed a new method of calculating the response novelty in AUT by looking at the term with the greatest semantic distance
- Verified that the new method more strongly correlated with human ratings versus currently existing metrics

SKILLS

Programming Languages: Python, R, SQL, Java, JavaScript, C, C++
Big Data & Machine Learning: Scikit-learn, PyTorch, NLTK, ggplot2, tidyverse, dplyr, numpy, pandas, scipy, matplotlib, SparkAPI
Data Science & Miscellaneous Technologies: A/B testing, Data science pipeline (cleansing, wrangling, visualization, modeling, interpretation), Tableau, Statistics, Time series, Experimental design, Hypothesis testing, APIs, Excel, Git, Beautiful Soup, Selenium, Scaled Agile, Jira Leankit, Azure DevOps, Databricks, NVIDIA NeMo, Praat

EDUCATION AND HONORS

The University of Texas at Austin, College of Natural Sciences May 2022
Bachelor of Science and Arts Honors in Computer Science Austin, TX
Certificates: Evidence and Inquiry
Polymathic Scholars Honors Program January 2019 – May 2022
Student in the Honors Program, Peer Mentor for Freshmen Polymaths

- Created a field of study called Computational Psychotherapy, focusing on the idea of language analysis in Psychotherapy
- Researched how sentiment and phonetic analysis could be used in a psychotherapy setting to understand a client better
- Thesis: Computational Methods in Psychotherapy: Prosodic and Semantic Analysis Framework for Clients with Major Depression (completed May 2022, https://repositories.lib.utexas.edu/handle/2152/114246)
- Mentored new students and community building through outreach with various seminars, socials, and discussions