Last Updated on 5th May 2022

Kris Smith

GitHub | LinkedIn | Kaggle
Kristopher.Smith@colorado.edu
21301 HWY 410 166 Bonney Lake WA, 98391

EDUCATION

UNIVERSITY OF COLORADO BOULDER
MS in Data Science
Graduating May 2023
Current GPA: 3.7 / 4.0

MIT - MITX
MicroMasters
Data Science
Sep 2020 - June 2022

IBM - COURSERA
Professional Certificate
Data Science
March 2020 - Sep 2020

COURSEWORK

GRADUATE
Machine Learning
Ethics In AI
Data Base Design
Algorithms
Data Mining
Cloud Computing
Big Data Architecture
Parallel Computing
Statistical Modeling

CERTIFICATE
Data Analysis
Probability and Statistics
Data Mining
Data Visualization
Deep Learning
Statistical Inference
Hypothesis Testing
Recommendation Systems
Computer Vision
Natural Language Processing

SKILLS

HARD
English • Python 3
R • Jupyter • LaTeX
HuggingFace • Excel
MySQL • Git • NoSQL

SOFT
Humility • Passion
Listening • Public Speaking
Integrity • Lifelong Learner

TECHNICAL PROJECTS

GOOGLE / JIGSAW | Toxic Comment Rater
Dec 2021 | Machine Learning Competition

- Developed a machine learning model to rate the severity of toxicity within comments.
- This was for an NLP machine learning competition hosted by Google's Jigsaw.
- Utilizing the **HuggingFace** platform, I employed transfer learning with the **Re-BERTa** base model, and blended with **XGBoost** and an ensemble of weak learning linear models including **Ridge**, **Lasso** and **Elastic Net**.
- Tools: **Python 3.8**, **Pandas**, **Numpy**, **Scikit-Learn**, **Jupyter**, **HuggingFace**

RESEARCH | Is Covid-19 Becoming Less Deadly?
Feb 2022 – Ongoing | Personal Project

- Created a system to model how lethality of Covid-19 has changed over time.
- Scraped reported data from the Johns Hopkins University database.
- Cleaned, transformed and analyzed the data with **R**.
- Visualisations with **ggplot2** and results compiled in **R Studio notebook**.
- Tools: **R**, **GitHub**, **Markdown**, **Tidyverse**, **Lubridate**, **R Studio**

GOOGLE / LIFECLEF | Bird Call Classifier
May 2022 | Machine Learning Competition

- Utilizing **Machine Learning** techniques to identify bird species by sound.
- This is a global competition on **Kaggle** hosted by Google, LifeCLEF and Cornell University with the goal of helping to advance the science of bioacoustics and support ongoing research to protect endangered Hawaiian birds.
- Trained a **Deep Learning classifier** which converts audio to mel-spectogram format, and with a **pre-trained ResNet** and **PyTorch** the spectogram images are then classified.
- Tools: **PyTorch**, **Python 3.8**, **Librosa**,**Jupyter**,**Pandas**,**Numpy**

EXPERIENCE

COUNCIL DATA PROJECT | Data Scientist
Dec 2021 – Ongoing | Seattle, WA

- Collaborate with other **volunteers** to build civic tech allowing for more transparency of various local governments.
- Contribute **ideas** and **code** for developing our **data API** so that data scientists/analysts and researchers can use our data more easily in their work.
- Motivating **analysis** of our data with our tools by building **Jupyter notebook** examples for future analysts/researchers.
- Building **machine learning models** for **auto transcription**, **speaker classification**, **named-entity-recognition**, and **sentiment analysis**.
- **Tools: HuggingFace**, **NoSQL**, **Scikit-Learn**, **Jupyter**, **Python 3**, **GCP(Firebase, Firestore)**