**JIM ZIELEMAN** | (571) 383-8584 | [zielemanj@gmail.com](mailto:zielemanj@gmail.com) | [www.zieleman.dev](http://www.zieleman.dev)| [github.com/LeafmanZ](https://github.com/LeafmanZ)

**EXPERIENCE**

**Department of Defense (DOD): GS-11** **Dahlgren, Virginia**

*Data Scientist – description is vague due to the sensitive nature of the work July 2021 – Current*

* Natural Language Processing (NLP) work involved: Named Entity Recognition (NER), Topic Modeling, Semantic Similarity, Sentiment Analysis, and Relation Extraction. Trained and fine-tuned SpaCy and HuggingFace NLP models using Python using a Pytorch backend. Many of my models reached a testing f-1 accuracy of over 95%.
* Combined NER models and topic modeling to enhance data science efforts. The NER model used a RoBERTa base. The topic modeling leveraged BERT embeddings and a class-based TF-IDF to create dense clusters. Then I used UMAP to reduce the dimensionality of embeddings and HDBSCAN to cluster reduced embeddings. This was adopted and integrated into the workflows of 2 divisions ~40 people.
* Built tools using Object Detection, and Optical Character Recognition (OCR) that were adopted and integrated into workflows of 2 divisions.
* Integrated all my deep learning models into fully functional end to end web applications using Python, Jinja, JavaScript, and HTML hosted on an internal server. These applications were built from scratch in a team of 2. My applications increased the speed of workflows for other divisions on average by ~200%.
* Used Git to manage the code base of projects, and Jira to manage Agile project workflows within my team.
* Trained models using cloud infrastructure like AWS Sagemaker and on internal servers and used Docker extensively for containerizing applications.
* Added powerful visualizations to applications I developed by using a combination of plotly, HTML, CSS, and JavaScript

**Department of Defense (DOD) & Joint Personnel Recovery Agency (JPRA)** **Remote**

*Applied Machine Learning Intern June 2020 – August 2020*

* Participated in a grant program from the DOD that was funded by Eccalon LLC. Developed applications for military and civilian use. My work was performed remotely in conjunction with the JPRA and the DOD.
* Built a fully functional cloud pipeline intended for applying deep learning tools to create models.
* Utilized a virtual machine to run scripts which scraped data, compiled it into a cloud storage bucket, and then transformed the data into an ingestible form for deep learning models.
* Extensively used GPU accelerated deep learning and some use of TPUs.
* Created an extensive plant database using SQL for the JPRA which housed important information about the plants and applied Google Earth to draw polygons needed to store geographic regions where the plants existed.
* Developed new plant image identification models using top image classifiers as a base. Performed research in other properties of plants that also required different deep learning techniques. The model was refined to a top 3 genus accuracy of ~99%.
* Developed the foundation of a Global Positioning System that did not require any connection.

**Madison Algorithmic Research:** **Harrisonburg, Virginia**

*NLP Engineer & Researcher Fall 2019 – Spring 2021*

* Performed NLP on 10K forms, NLP on news feeds, and aggregating different sentiment generators to forecast stock prices.
* Wrote VBA & Python functions that calculated Option Price, Delta, and Vega using a Black Scholes and a binomial tree implementation allowing for either or both continuous dividends and discrete dividends. The binomial tree calculation also factored in early exercise and with altered parameters would be able to calculate implied volatility. Included arguments to calculate either European or American options.
* Created an automated multi-Greek option hedging timeline which would use predicted volatility to determine profitable positions.

**EDUCATION**

**­­­ University of Virginia** **Charlottesville, Virginia**

*Master of Science in Data Science*

*Expected Graduation Date - May 2024*

* **Cumulative GPA:** N/A
* **QFIN Coursework:** Programming for Data Science, Linear Models for Data Science, Practice and Application of Data Science, Statistical Learning, Foundations of Computer Science, Bayesian Machine Learning, Deep Learning, Ethics of Big Data

**James Madison University** **Harrisonburg, Virginia**

*Bachelor of Science in Quantitative Finance with a concentration in Risk Management*

*Bachelor of Science in Mathematics*

* **Cumulative GPA:** 3.543
* **QFIN Coursework:** Securities Pricing, Elemental and Derivative Securities Analysis, Financial Modeling and Risk Analysis, Risk Management, Financial Accounting, Corporate Valuation, Microeconomics, Macroeconomics, Mathematical Finance, Financial Data Science
* **MATH Coursework:** Computers & Numerical Algorithms, Topology, Real Number Systems, Real Analysis, Abstract Algebra, Probability and Statistics, Linear Algebra, Differential Equations, Multivariate Calculus, Discrete Mathematics

**PROJECTS**

**NLP Python Library Developer: Spacy-Annotator** **Remote**

[*https://github.com/ieriii/spacy-annotator*](https://github.com/ieriii/spacy-annotator) *&* [*https://pypi.org/project/spacy-annotator/*](https://pypi.org/project/spacy-annotator/) *October 26, 2021 - Current*

* Second largest contributor (second only to founder) to spacy-annotator an open source python library that annotates text data for Named Entity Recognition (A subfield of NLP). The library receives over 1,000 downloads a month (https://pypistats.org/packages/spacy-annotator).
* Developed functionality to convert legacy dataframe annotations into a modern spac3y (.spacy encodings) ingestible format.

**Personal Portfolio Website:** [**www.zieleman.dev/**](http://www.zieleman.dev/) **Remote**

*Full Stack + Leveraging Cloud Infrastructure January 5, 2021 - Current*

* Used Python, Flask, HMTL, JavaScript, Jinja, CSS, Bootstrap to build out my personal portfolio website.
* Used Docker to containerize my website and deployed it to AWS Lightsail.
* My website hosts papers I wrote with interactive graphs, financial calculators implemented from academic papers, offers option pricing calculator code in python and VBA, certifications, a further detailed resume, and my personal data science projects.

**Personal Portfolio NLP Named Entity Recognition and Relation Application:** [**www.zieleman.xyz:5000**](http://www.zieleman.xyz:5000/) **Remote**

*Full Stack + Leveraging Local Infrastructure + SOTA Heavy AI/ML March 15, 2021 - Current*

* Built a State of the Art (SOTA) Natural Language Processing web application that is served off local infrastructure.
* This application prioritizes documents for a user to read by identifying relevant entities within a document and dividing the number of entities by word count respectively. This generates a document entity richness score that can be used as our prioritization metric.
* The application offers a large selection of common entities that can be detected to allow users to select what entities are important to them.
* For multiple users the application has functionality allows for even distributions of document reading workload to each user.
* Larger sets of documents are clustered into topics that are non-deterministic and are identified by a set of key themes/words. In the case of multiple readers these sets of clusters will be assigned evenly to enable faster development of domain knowledge.
* Documents and entities will be additionally related to each other inside a network graph for users to visually examine the relationships.

**Personal Portfolio Object Detection and Optical Character Recognition Application:** [**www.zieleman.xyz:5010**](http://www.zieleman.xyz:5010) **Remote**

*Full Stack + Leveraging Local Infrastructure + SOTA Heavy AI/ML April 20, 2021 - Current*

* Built a State of the Art (SOTA) application that extracts text from images and/or photocopies of documents.
* The application detects if a document is tilted or slanted and rotates the document against the angle of the tilt detected.
* The application applies homography to the document to flatten out warped and curved distortions within a document.
* After straightening out the document I used a modified version of CRAFT (object detection) to identify all the bounding boxes for each line of text.
* To convert the text detected into raw text I chipped out the bounding boxes from CRAFT and used a vision transformer (ViT) based OCR processor to extract the text.

**Devpost: ClimateHacks Hackathon:** [**https://devpost.com/software/the-value-of-forest**](https://devpost.com/software/the-value-of-forest) **Remote Work, Virginia**

*Data Scientist April 23, 2021 - April 25, 2021*

* Used Python and JavaScript on Google Earth Engine to build a carbon stock valuation application for a hackathon that worked on climate saving applications.
* The carbon stock evaluation model correlated the normalized difference in vegetation index with the WHRC global above ground biomass dataset to come up with the regression formula.
* Used the Filmora video editing software to create the pitch video of the application. Won an honorable mention.
* Developed an algorithm that valuated the carbon stock value of a selected region on google maps with the current market price of carbon.

**TECHNOLOGY STACK**

**Full Stack:** HTML, JavaScript, CSS, DataTables, Bootstrap, Python, Flask, Gevent, WSGIServer, AWS (Lightsail, Route53, EC2/3, S3), GCP, Jinja.

**Data Science:** Transformers, Convolutional Neural Networks (CNN), Natural Language Processing (NLP), Named Entity Recognition (NER), Semantic Similarity, Topic Modeling, Data Visualization, Greedy Algorithm, Pandas, BERT, RoBERTa, UMAP, TF-IDF, HDBSCAN, Pandas, NetworkX, HuggingFace, SpaCy, Pytorch, Sentence-Transformers, SKlearn, Cupy, Plotly. Numpy, Computer Vision, Vision Transformers (ViT), Object Detection, Object Classification, Optical Character Recognition (OCR), SQL, REST, MS Excel, SPSS, SAS, VBA, Python, AWS Sagemaker.

**Coursera Certifications:** Deep Learning Specialization, Convolutional Neural Networks, Sequence Models, Structuring ML Projects, Improving DNNs: Hyperparameter Tuning, Regularization, and Optimization, NNs & DL, GCP Fundamentals, Building Streaming Analytics Systems on GCP, Data Engineering + Big Data + ML on GCP, Smart Analytics + ML + AI on GCP, Building Batch Data Pipelines on GCP, Modernizing Data Lakes and Warehouses on GCP.

**Additional Certifications:** Bloomberg Market Concepts, Udacity AI for Trading.

**­**