ERIC BOXER

ecboxer.github.io

Email : ecboxer@gmail.com
Mobile : (347) 525-3516

Professional Experience

- **Swiss Re** *Data Scientist* Jan 2020 – Present

  Early Warning Signals: Developed a Python package to extract "signals" from an ElasticSearch database of news articles. Viewing the corpus as a network (with keywords as nodes and co-occurrence of keywords in an article forming edges), we extract communities from the network and surface important articles for internal users in a dashboard. Presented the method at AAAI and internal/external industry events to technical and non-technical audiences.

  Sentiment: Developed repeatable aspect-based sentiment analysis (ABSA) approach for topics such as COVID-19 medical developments and client news. By either gathering a set of articles with "noisy sentiment labels" (based expert-driven queries) or starting with an open-source pre-trained domain-relevant model, we finetuned ABSA models for our task. Model output was presented in a dashboard to medical experts and analysts to curate their news consumption.

  Alternative Data Sources: Trained interpretable models to predict underwriting outcomes using traditional life insurance data (e.g. application responses, prescription histories, driving records, past insurance history) and new 3rd-party data sources (e.g. electronic health records, billing records, clinical labs), to assess the ability of the new data sources to replace the traditional sources, for multiple clients.

  Accelerated Underwriting: Trained an interpretable model to reduce referrals by 30%, for a mid-sized US life insurer, thereby reducing costs. Delivered applicant cohorts (created with unsupervised self-organizing maps) and insights about the client's underwriting process that were incorporated into their rules going forward.

  Excel Extraction: Trained models to extract variable-format tabular data from Excel Statement of Value (SoV) documents, containing information about properties in insurance submissions. The approach borrows from NLP techniques (namely, named entity recognition) and involves 5 independent models feeding into a metamodel which provides final predictions. Also, trained models to normalize free-text values (i.e. multi-class classification). The pipeline provides input to a submission triage model and is planned to gradually replace the current manual entry of SoVs.

  Culture: Organized monthly meetups on data science-related projects for Swiss Re colleagues. Reached out to potential speakers throughout the company and externally, hosted the meetups and wrote up summaries for internal publication.

- **Swiss Re** *Data Science Intern* May 2019 – Dec 2019

  Early Warning Signals: Developed a novel outlier detection method for news articles, to identify potentially critical information for decision makers. Viewing the corpus as a network, we are able to identify important articles as those which form links between otherwise distant terms. Also, extended the node2vec method of computing node embeddings to include domain knowledge (by warm-starting node embeddings) and early stopping (by computing link prediction scores on a hold-out validation network).

  Property & Casualty Solutions: With financial and geospatial data from the client/insurer's portfolio, explored drivers of named-event (i.e. caused by significant events such as fires or floods) and attritional loss. Integrated in-house natural catastrophe models and publicly available fire incident data to create and validate policy- and broker-level risk scores. Recommendations were presented to the client, who expressed interest in continued engagement with the analytics team and reinsurance renewal with Swiss Re.

Education

- **Columbia University** New York, NY
  MS in Data Science; GPA: 3.8 / 4.0 *Aug 2018 – Dec 2019*
  Capstone Lymphoma Development Prediction: In a large EHR database, matched similar patients who went on to be diagnosed with diffuse large B-cell lymphoma (DLBCL) and those who did not. Applied NLP techniques to create patient representations, cluster patients and predict DLBCL diagnoses.

- **SUNY Stony Brook** Stony Brook, NY
  BS in Mathematics and Economics, Minor in Philosophy; GPA: 3.9 / 4.0; Phi Beta Kappa *Aug 2013 – Dec 2016*

Publications

- N. Le Vine, **E. Boxer**, M. Dinani, P. Tortora, and S. Das. Identifying early warning signals from news using network community detection. In *AAAI Conference on Innovative Applications of Artificial Intelligence*, 2022.

Programming Skills

- **Python**: PyTorch, Keras, Scikit-Learn, spaCy
- AWS, Azure, D3, Elasticsearch, Emacs, Git, Javascript, Jupyter, LaTeX, Linux, R, SQL, VSCode