| Avoid Topics | Possible Topics |
| --- | --- |
| Health Stuff | Stock Market Prediction |
| Anything computer vision NOOOO :p | Climate change |
| No images | Fake News Detection |
|  | AI for Good |
| Wala emotion wala personality | NLP |
|  |  |

Team Name:

Shortlisted:

1. Crypto - Forex - **Stocks**: Dany
2. Disaster relief: hamza, responding to the disaster in real time,
   1. Joplin 2011:206,764 tweets collected during the tornadothat struck Joplin, Missouri (USA) on May 22, 2011. Re-searchers at the University of Colorado at Boulder collectedthe dataset through Twitter’s API using the hashtag1#joplin.
   2. Sandy 2012:140,000 tweets collected during the Hurri-cane Sandy, that hit Northeastern US on Oct 29, 2012. Thedataset was collected using the hashtags#sandy,#nyc.
   3. We can also collect data about beirut explosion if easy
   4. [During emergencies, this information is made widely available, at no cost, through mechanisms such as Humanitarian Information Centres (HICs), Reliefweb, AlertNet and the UNOSAT web portal.](https://odihpn.org/magazine/using-satellite-imagery-to-improve-emergency-relief/)
   5. [Using imagery to best effect in disaster relief – MapAction](https://mapaction.org/using-imagery-to-best-effect-in-disaster-relief/)
   6. [2020 CES: How Intel AI Helps the Red Cross Generate Open Source Maps for Disaster Relief](https://newsroom.intel.com/news/how-intel-ai-helps-red-cross-generate-open-source-maps-disaster-relief/)

**How we may use twitter in real-time classification of disaster damage? :**

<https://medium.com/@udayshankarmenon/how-we-may-use-twitter-in-real-time-classification-of-disaster-damage-part-1-ed5e3bd6889d>

1. Modeling demand (transportation and traffic) : hiba

Comparison of short-term traffic demand prediction methods for transport services: <https://www.research-collection.ethz.ch/bitstream/handle/20.500.11850/356143/ab1447.pdf?sequence=1&isAllowed=y>

Travel demand modelling, data collection and well-being: <https://link.springer.com/article/10.1007/s11116-019-09987-5>

Demand Forecasting in Transport: Overview and Modeling Advances:

[**https://www.tandfonline.com/doi/pdf/10.1080/1331677X.2011.11517446#:~:text=Transport%20demand%20forecasting%20models%20can,and%20(d)%20traffic%20assignment**](https://www.tandfonline.com/doi/pdf/10.1080/1331677X.2011.11517446#:~:text=Transport%20demand%20forecasting%20models%20can,and%20(d)%20traffic%20assignment)**.**

**An agent-based modeling system for travel demand simulation for hurricane evacuation**

[**https://www.sciencedirect.com/science/article/abs/pii/S0968090X14000576**](https://www.sciencedirect.com/science/article/abs/pii/S0968090X14000576)

**Available dataset: Three repeated cross-sectional household travel survey datasets collected in the greater Toronto and Hamilton area (GTHA) are used for the investigation.**

[**https://tspace.library.utoronto.ca/handle/1807/76810**](https://tspace.library.utoronto.ca/handle/1807/76810)

**Many research papers have been applied so data shouldnt be a problem.**

Fintech:

1. Crypto predictions with respect to highly influential tweets

Social:

Links :

1. [Microsoft AI for Earth](https://www.microsoft.com/en-us/ai/ai-for-earth)
2. [AI for Social Good: 7 Inspiring Examples](https://www.springboard.com/blog/ai-for-good/)
3. [Here are 10 ways AI could help fight climate change](https://www.technologyreview.com/2019/06/20/134864/ai-climate-change-machine-learning/)
4. [Shady’s publications](https://scholar.google.com/citations?hl=en&user=Fe8iIHMAAAAJ&view_op=list_works&sortby=pubdate)

A friend of hamza has sho ismo: solar panel, we will have meeting to suggest what do?

Shady’s Roadmap:

1. first month:
   1. Find data
   2. Clean data
   3. Read literature
2. Second month:
   1. Experiment models
   2. Choose models from literature

<https://www.kaggle.com/c/jane-street-market-prediction/overview>

1. Stock market prediction

In the first three months of this challenge, you will build your own quantitative trading model to maximize returns using market data from a major global stock exchange. Next, you’ll test the predictiveness of your models against future market returns and receive feedback on the leaderboard.

Your challenge will be to use the historical data, mathematical tools, and technological tools at your disposal to create a model that gets as close to certainty as possible. You will be presented with a number of potential trading opportunities, which your model must choose whether to accept or reject.

In general, if one is able to generate a highly predictive model which selects the right trades to execute, they would also be playing an important role in sending the market signals that push prices closer to “fair” values. That is, a better model will mean the market will be more efficient going forward. However, developing good models will be challenging for many reasons, including a very low signal-to-noise ratio, potential redundancy, strong feature correlation, and difficulty of coming up with a proper mathematical formulation.

<https://www.kaggle.com/c/data-science-for-good-city-of-los-angeles>

Data Science for Good: City of Los Angeles

Help the City of Los Angeles to structure and analyze its job descriptions

The City of Los Angeles faces a big hiring challenge: 1/3 of its 50,000 workers are eligible to retire by July of 2020. The city has partnered with Kaggle to create a competition to improve the job bulletins that will fill all those open positions.

Problem Statement

The content, tone, and format of job bulletins can influence the quality of the applicant pool. Overly-specific job requirements may discourage diversity. The Los Angeles Mayor’s Office wants to reimagine the city’s job bulletins by using text analysis to identify needed improvements.

The goal is to convert a folder full of plain-text job postings into a single structured CSV file and then to use this data to: (1) identify language that can negatively bias the pool of applicants; (2) improve the diversity and quality of the applicant pool; and/or (3) make it easier to determine which promotions are available to employees in each job class.

From <<https://www.kaggle.com/c/data-science-for-good-city-of-los-angeles>>

<https://arxiv.org/pdf/1906.05433.pdf>

Collaboration makes it easier to develop effective strategies. Working with domain experts reduces the

chance of using powerful tools when simple tools will do the job, of working on a problem that isn’t actually

relevant to practitioners, of overly simplifying a complex issue, or of failing to anticipate risks.

Modeling demand:

Modeling demand and planning new infrastructure can significantly shape how long trips are and which transport modes are chosen by passengers and shippers – for example, discouraging sprawl and creating new transportation links can both reduce GHG emissions. ML can provide information about mobility patterns, which is directly necessary for agent-based travel demand models, one of the main transport planning tools [173]. For example, ML makes it possible to estimate origin-destination demand from traffic counts [174], and it offers new methods for spatio-temporal road traffic forecasting – which do not always outperform other statistical methods [175] but may transfer well between areas [176]. Also, short-term forecasting of public transit ridership can improve with ML; see for example [177, 178]. ML is particularly relevant for deducing information from novel data – for example, learning about the behavior of public transit users from smart card data [179, 180]. Also, mobile phone sensors provide new means to understand personal travel demand and the urban topology, such as walking route choices [181]. Similarly, ML-based modeling of demand can help mitigate climate change by improving operational efficiency of modes that emit significant CO2, such as aviation. ML can help predict runway demand and aircraft taxi time in order to reduce the excess fuel burned in the air and on the ground due to congestion in airports.

The availability of large quantities of data, combined with affordable cloud-based storage and computing, indicates that industry may be an excellent place for ML to make a positive climate impact.

disDisaster response High Leverage

In disaster preparation and response, two types of ML tasks have proven useful: creating maps from aerial imagery and performing information retrieval on social media data. Accurate and well-annotated maps can inform evacuation planning, retrofitting campaigns, and delivery of relief [635, 636]. Further, this imagery can assist damage assessment, by comparing scenes immediately pre- and post-disaster [637, 638]. Social media data can contain kernels of insight – places without water, clinics without supplies – which can inform relief efforts. ML can help properly surface these insights, compressing large volumes of social media data into the key takeaways, which can be acted upon by disaster managers [624, 639, 640].

**A Machine Learning Approach to Address Methane Emissions from Oil and Gas Operations**

Hide abstract

Methane emissions mitigation in the oil and gas sector is an important component of climate policy across North America. One of the common features of methane regulations is the implementation of periodic leak detection and repair (LDAR) surveys. These surveys are typically conducted by a two-person crew visiting each site with a handheld sensor to detect methane leaks. However, this is time-consuming and expensive, particularly for companies that manage thousands of production sites. Furthermore, recent research has shown that emissions follow a skewed distribution where a small number of sites (‘super emitters) are responsible for a majority of total methane emissions. Thus, finding these super-emitters in a timely manner will optimize both time and financial investments in emissions reduction programs. Here, we propose to develop a machine learning model to predict high-emitting sites that can be prioritized for LDAR surveys. Such prioritization can reduce the cost of surveys and increase emissions reductions compared to a random survey approach currently used in the industry. Prior efforts have failed because of limited data availability. By taking advantage of recent measurements across North America, we will assemble the largest complete data-set on methane emissions to predict the occurrence of super-emitters. Through this mentorship program, we look forward to discussions on the role of machine learning in addressing emissions reduction policies. Specifically, we are looking for 1) feedback on research design, including defining potential attributes for predictive models and 2) suggestions on improving model performance.

**Scraping tweets:**

<https://analyticsindiamag.com/complete-tutorial-on-twint-twitter-scraping-without-twitters-api/>

One of the biggest challenge during this project was acquiring post-disaster tweets. Twitter offers a very useful API to obtain tweets, however, there were difficulties specific to our objective:

* A cap on the maximum number of tweets scraped. For the model developed, it would have taken few weeks to generate the data.
* Location of tweets is often turned off by users — this prohibited geographical mapping of ground damage. This task may require more expertise and information.
* Collecting tweets from historical events is challenging.

One particular data source that helped us was [CrisisLex26](https://github.com/sajao/CrisisLex/tree/master/data/CrisisLexT26/), a suite of nearly ~250K tweets from 26 different disasters between 2012 and 2013. This is an excellent source of information labeled by crowdsource workers according to informativeness (informative or not informative), information types (e.g. caution and advice, infrastructure damage), and information sources (e.g. governments, NGOs). [CrisisLex.org](https://crisislex.org/) is a repository of similar crisis related social media data and tools.

For this task, 6260 California wildfire tweets were scraped using Tweepy, of which, more than half were duplicates unfortunately. This goes to justify our use of open source historical tweets. Our multi-classifier model gave the following distribution of classes:

The project repository is available [here](https://github.com/ukm5/ClientProject_NewLightsTechnologies).

NLP project basically !