STATEMENT OF WORK

_____

A practical fake news detection: Fake news and disinformation are increasingly spreading on the internet, especially on social media. How to stop disinformation infiltrating into daily life? A well-trained machine learning module can help!

_____

THIS STATEMENT OF WORK (this "**SOW**") is entered into by and between the undersigned Buyer and Service Provider pursuant to the Project Services Agreement (the "**Agreement**") governing the following Project:

Elance Project Name: Dragnet – Fake News Detector
Elance Project Team: Hung-Yi Chen 100845864

All capitalized terms not defined in this SOW have the meanings given to such terms in the Agreement, unless the context requires otherwise. The information in this SOW shall be considered Confidential Information under the terms of the Agreement.

1 Project Description

According to MIT researchers in 2018, "Rumors can spread faster and farther than the truth." As an enormous news got published every day, building up a machine learning module for detecting fake news to make sure malevolence will be prevented promptly.

2 Key Assumptions

Dataset should have at least five columns:
1. news ID: unique ID for each peace of news
2. news title
3. author: full name of the article author
4. news text: news content
5. fake news label: 0 stands for non-fake, 1 stands for fake (test dataset should be without label)

Dataset limitation and constraints:
1. ideally have one balance dataset, two unbalance datasets (one with higher percentage of fake news, the other with higher percentage of non-fake news)
2. news published time should be after 2017 (prevent 2016 U.S. president election)

3 Scope of Services

This ML module will use news ID, news title, author, news text, and prelabeled column to train the module. The dataset will be from Kaggle.com, database should be after 2017 (prevent 2016 U.S. president election)

4 Key Test Indicators

1. Prediction accuracy should be over 80% in test data
2. Module can be used in a real world and reach prediction rate to 60%
3. Processing time should not over 5 minutes