DATA SCIENCE CAPSTONE PROJECT

Presented by Dennis Lam, 21 March 2021

ABOUT THIS PROJECT

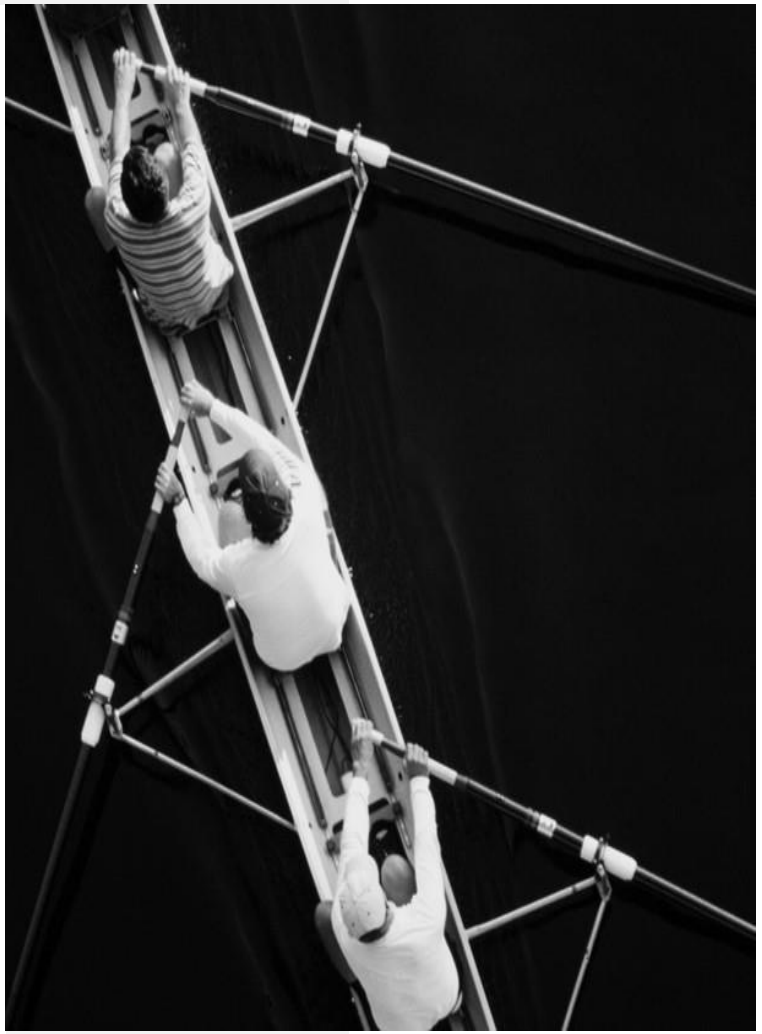
This project is a natural language processing where an application is developed using NLP algorithms to predict next word that is typed by the user in a phone or computer.

Project Deliverables:

- An NLP model.
- Web based data product.
- Presentation.

MODEL CREATION STEPS

- **Understanding the problem** – Need to help users predict next word
- **Data acquisition and cleaning** – Blog US text is selected and used NLTK library for text preprocessing
- **Exploratory analysis** – Unearth trends and patterns in text
- **Statistical modeling** – A basic N-gram model is built
- **Predictive modeling** – Unigrams, Bigrams and Trigrams models are built and tested for predictions
- **Creative exploration** - Deep learning model is preferred due to higher accuracy and use less memory
- **Creating a data product** – Use Streamlit to run the application online
- **Creating a presentation pitching your product**

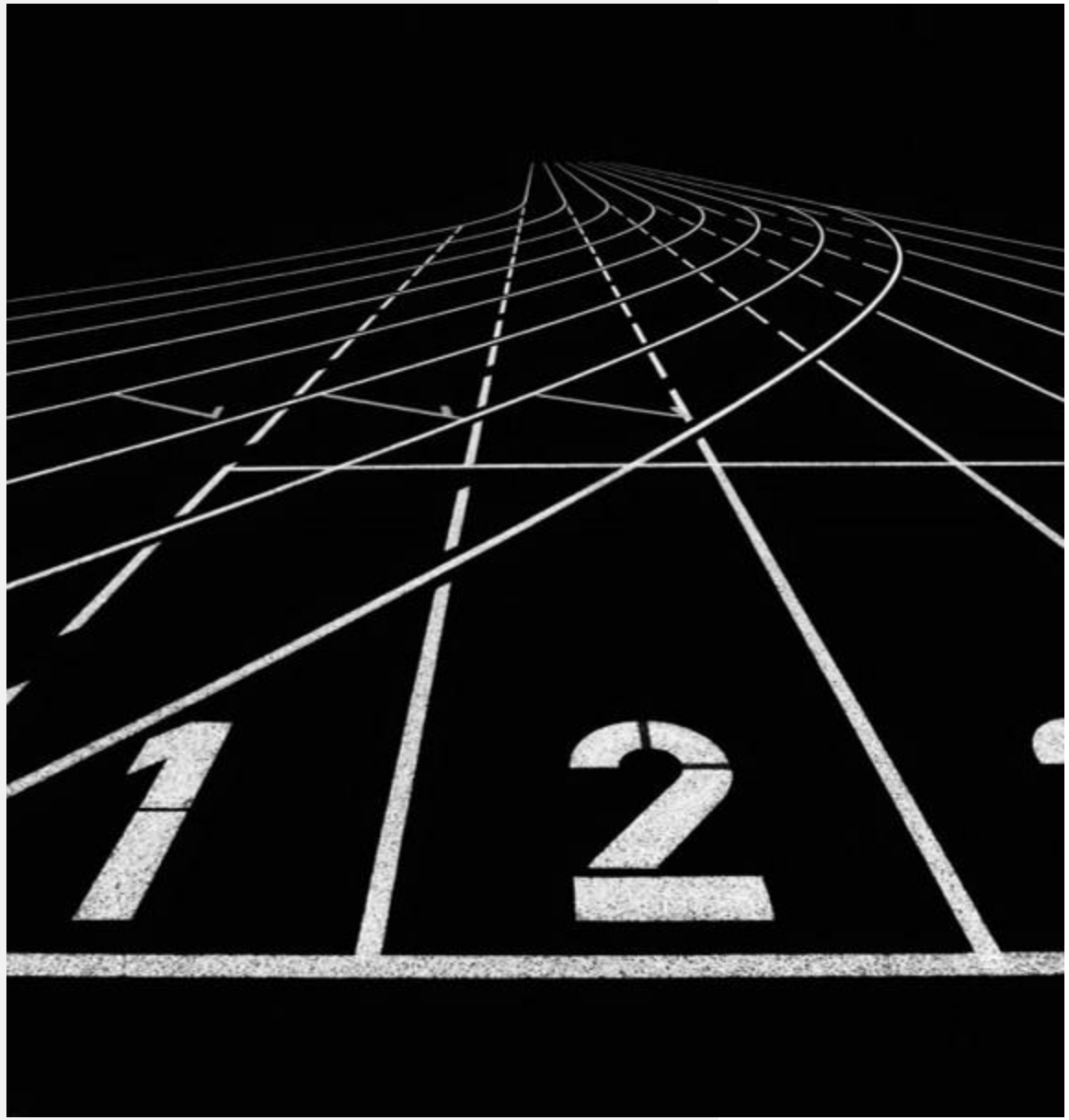


APPLICATION FEATURES

- *Takes an input phrases to predict next word.*

- *A slider for user to output minimum 1 to maximum 10 suggested words*

- *Smooth response and accurate*



REFERENCES CITED AND NLP PRACTICE

1. Language Modeling With NLTK (https://medium.com/swlh/language-modelling-with-nltk-20eac7e70853)
2. A Comprehensive Guide to Build your own Language Model in Python!(https://www.analyticsvidhya.com/blog/2019/08/comprehensive-guide-language-model-nlp-python-code/)
3. Sequence Models (https://www.coursera.org/learn/nlp-sequence-models?specialization=deep-learning)
4. Natural Language Processing with Probabilistic Models (https://www.coursera.org/learn/probabilistic-models-in-nlp?specialization=natural-language-processing)
5. Natural Language Processing in TensorFlow (https://www.coursera.org/learn/natural-language-processing-tensorflow?specialization=tensorflow-in-practice)
6. Applied Text Mining in Python (https://www.coursera.org/learn/python-text-mining?specialization=data-science-python)