Hello, there! My name is Diego, and I am part of the Master of Software Development 2022 cohort at the University of Utah.

The application I have developed for my capstone project is called Distance Arbitrage aka “distarb”. A bit of background for what arbitrage is for those who might not know:

Slide 2

Arbitrage is taking advantage of the difference in prices between markets. A historical example would be a speculator with a strong telescope could read the communication flags of an incoming ship that was carrying grain before their competitor standing next to them atop the hill with weaker telescope.

The speculator with the stronger telescope could then know before their competitor that the incoming shipment of grain had spoiled and could sell their shares of the incoming shipment and make money, leaving their competitor holding the bag.

Slide 3

What my program does is that it looks for mistakes humans might make when performing a transaction. It does so using the edit distance, in other words, the adds and deletes one must make to modify a word to change it into another word.

For example: cat to car has an edit distance of one. Because the only the t is changed to an r.

Slide 4

But cat to cars would be two because there is an insertion of the letter s in addition to the substitution

Slide 5

What we all (presumably) learned in grade school is how to find the physical distance between two points thanks to our good friends Euclid and Pythagoras. Distarb also calculates the physical distance between keys.

Slide 6

(PAUSE)

Well, sort of.

Slide 7

This is a generic and not very interesting problem if keyboards were in neat little rows and columns. The diagonal distance between two keys offset by one from each other is the same no matter what two consecutive rows you look at. But keep an eye on the left-most diagonal line from S to C and how it relates in length the diagonal line in the top right from u to h.

Got it? Good.

Slide 8

As you can see, the linear distance from the S to the C is significantly longer than the linear distance from U to H. this is because the offsets on a standard keyboard, and I would wager the offsets on your keyboard right now are 1/2 key, 1/4 key and 1/2 key. This results in the diagonal distance being varied between rows. I believe the varied distance changes the chance of a key being pressed by mistake.

Slide 9

My idea is to use the edit distance of words and physical distance of keys in an application where the user can enter in the name of a company that is about have its initial public offering or any currently publicly traded company and the program will return a list of the tickers or companies a person is most likely to incorrectly enter, due to fat fingering, or careless reading.

Slide 10

My inspiration for this came years ago when Facebook had its initial public offering and a little known Spanish company with the ticker FACE saw an uptick in its share price because people confused it for Facebook, which IPO’d as FB. As you may well be aware, Facebook is not a Spanish company specialising in biometric authentication. But people were confused and mistakenly purchased the wrong stock, in part due to seeing the similar name and that it was trading for significantly cheaper than Facebook. People jumped at the opportunity to own a piece of Facebook for a steal of a price. As the presentation subtitle says…

Slide 11

Pobody’s Nerfect!

Slide 12

Distarb also operates as a web scraper. It goes to the SEC’s website and pulls all the data of what stocks, bonds, etc. that US Senators are buying/selling and then gives details about the who, what, and when regarding the most recent stock transaction. Since the entire page is scraped, and the program is only returning the most recent stock purchase, there is a lot of information that can be used for analysis later on.

I know this has been a lot of talk about what it does. But a lot of this app is based on what is going on the backend.

Slide 13

Take a moment to familiarise yourself with the layout of the internals, and then let’s take a look at distarb!

Distarb in action

INSERT

Slide 14

To run, simply click on the app!

Distarb is written in Python3. Some of the notable libraries are:

Clavier - a library with tools for finding distances

Pandas

Pendulum for managing dates and times

PyQt5 for the graphical user interface. I chose PyQt5 because it is built right on top of C++. So you get the ease of use of python, and the speed of a lower level language.

And to manage all these packages, I used Anaconda. Anaconda is also used to create environments that can be accessed online by users who are having issues with setting up their environment on their local machine

Slide 15

The user can use a convenient dropdown interface to select either a ticker or a company’s listed legal name and display their stock prices from the last year.

Slide 16

Do a stock

In window to the right, the nearest matches based on edit and linear distance are listed.

In the graph, the stock price from the last year of the selection and the two most similar from the list are plotted.

I’ll start by selecting a ticker

Slide 17

Do a company

Companies are slower because of the algorithm having to loop through longer strings to find the nearest matches.

Slide 18

Enter a company

What we colloquially call a company is often different from its legal name. Apple, for example is legally Apple Inc. But, we all just call it apple. The search box at the bottom allows the user to enter custom searches for this reason. And it will return the closest tickers – like from the Facebook and FACE example in slide 10

Slide 19

Scrape

The scraper displays the most recent stock transaction listed on the SEC site. The graph of the stock price starts four weeks before the senator reportedly made the transaction and finishes today. The blue line indicates the transaction date.

Thank you so much for your time, and congratulations on making it through MSD!!