**Project Proposal**

**Colleen Dunlap**

**Background and Motivation**

I wanted to choose a project that was at the intersection of data analytics and social impact. Therefore, designing an ANN using the Missing Migrants dataset was a perfect fit. Immigration is a huge issue, both in countries people are immigrating to and emigrating from. In addition, migration is dangerous. I learned about the pain and struggles migrants deal with this spring when I visited El Paso and stayed at Casa Vides shelter for homeless migrants. The stories I heard shocked me. I knew this was an issue I couldn’t ignore.

**Problem Statement**

Every year, people flee their homeland to escape war, poverty, violence, and to seek a better life. In 2017, the number of migrants worldwide reached 258 million1 Too often, these individuals go missing and die, never reaching their destination. Before choosing this problem, I researched the effects and frequency of international migration, using resources from the Missing Migrants Project and the United Nations (all sources cited in the Bibliography).

**Choice of Dataset**

The Missing Migrants Project began with the October 2013 tragedy, when 368 people died in two shipwrecks off the coast of Italy2. They track the deaths of migrants who’ve gone missing en route and maintain a dataset of these events. Features for each event include the number of deaths, number of survivors, date of events, number of males, females, and children, and geographical data regarding the incidents3.

The dataset can be found here: <https://www.kaggle.com/snocco/missing-migrants-project>

**Goals of Model**

I want to use this dataset, which includes data on incidents occurring between January 2014-March 2019 to predict the number of expected tragedies and their location to help authorities and humanitarian groups predict these tragedies and intervene before crisis occurs. For example, a user can input the month and year they want to predict for, and they’ll be told what routes will be the most traveled and how likely a crisis is to occur, and the expected toll. This can help humanitarian workers better mobilize to help these at-risk migrants. In addition, this model can reflect the differences in death tolls caused by policy changes, seasonal conditions, and local violence.

**Tools**

For this project, I will be using the Python 3 language. I will be training my model with Keras packages, which utilize Tensorflow on the backend. I will be using Pandas data frames and NumPy to perform linear algebra/matrix operations on the data. I currently plan to use Matplotlib to visualize the data, however, I will also be researching other python plotting and visualization packages. I will use Methodologies such as feature scaling to improve model performance, and will implement a k-cross validation to validate the model on separate training and testing data.

**Bibliography**

1.United Nations, *International Migration Report.* Page 4. <https://www.un.org/en/development/desa/population/migration/publications/migrationreport/docs/MigrationReport2017_Highlights.pdf>

2. Missing Migrants. *About.* <https://missingmigrants.iom.int/about>

3. Missing Migrants. *Methodology.* <https://missingmigrants.iom.int/methodology>