**Data Mining and Machine Learning Project**

**Pavel Golikovski 320636574**

**Natan Zaitzev 312171705**

**Alla Davydov 317000669**

**Process of the Project Study**

We started out by reviewing data mining and machine learning algorithms related to our project

* KNN, SVM, Decision Tree
* RNN, LSTM and Bidirectional LSTM
* CNN
* Neural network layers
* Word embeddings

**Project Flow**

Our project consisted of mostly-working written code which attempted to solve the problem of detecting fake reviews on online review platforms using deep learning architectures:

1. Bidirectional LSTM network using GloVe 50d
2. Bidirectional LSTM network with an Attention layer using GloVe 100d
3. CNN + LSTM using Doc2Vec + TF-IDF

We had to split our reviewing process into multiple parts:

* Gathered all the required files (e.g., GloVe and datasets)
* Got the code running after cross-matching many different python libraries
* Decided to dedicate our efforts learning word embedding techniques first followed by the neural network models respectively

As we studied the theory behind these concepts, we used many different online resources alongside the course materials, we started to work on the presentation explaining how the models attempt to solve the aforementioned problem.

**Results and Conclusions**

* Working on the project exposed us to many more complex aspects, theories and ideas behind machine learning in general
* The project helped us to solidify what we had learned throughout the semester and associate between theory and machine learning in practice
* We realized that having a larger dataset could potentially increase the efficacy of neural networks, which we didn’t have
* We started with a rather simple neural network and learned that adding more complex and deeper layers does generally result in better predictions,

For example:

1. Using plain Bi-LSTM models yields worse result than more sophisticated models that include more layers such as an Attention layer
2. Chaining neural networks (CNN LSTM) can improve the model predictions

**References**

1. <https://github.com/ashishsalunkhe/DeepSpamReview-Detection-of-Fake-Reviews-on-Online-Review-Platforms-using-DeepLearning-Architectures>
2. <https://www.kaggle.com/rtatman/deceptive-opinion-spam-corpus>
3. <https://www.kaggle.com/watts2/glove6b50dtxt>
4. <https://www.kaggle.com/danielwillgeorge/glove6b100dtxt>