Reviews Moderation with Machine Learning

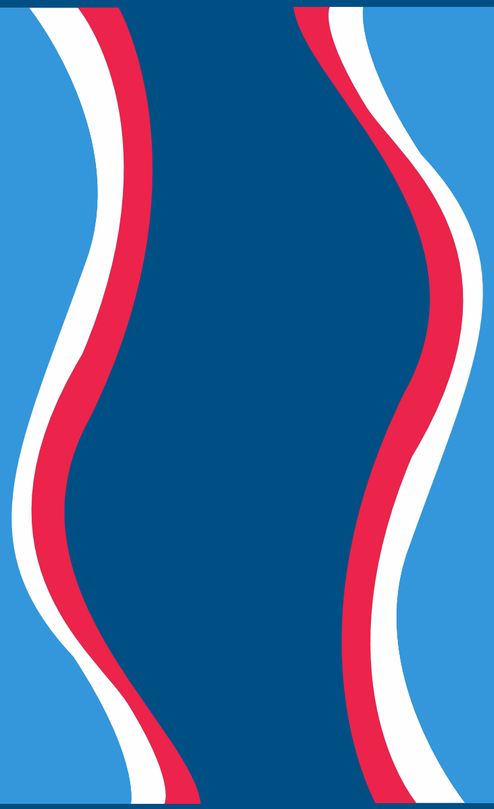
Business Problem

- Receiving approximately 500 reviews on a daily basis
- Manual review moderation

So...

Manual Review Moderation Problems

A lot of resources and time are being consumed from the customer service



Customers do not know when their review will be uploaded

Ideal Solution

Are there are instances that an evaluation may not be published?

We publish all the evaluations! You have to be sure that you have followed the following guidelines and you have read the terms of use.

- Any comment that is offensive or profane is not published.
- Personal data: we protect the privacy of your personal data and for this reason we erase any kind of personal data. You will never see published: surname, email, phone number etc.
- Prices & receipts: of course you can express your opinion upon the prices that a doctor offers for his service but doctoranytime is not going to publish details upon a financial transaction. doctoranytime is not a control mechanism and does not have the authority for doing so. After an appointment you have every right to ask for the receipt.
- Accuracy of diagnosis or treatment: It is extremely important to share your feelings on how the doctor treated your case. But we can not publish comments whether the accuracy of the diagnosis or treatment. In order to be able to prove if the doctor was right or wrong personal data of your case should be published and this is prohibited by law. Those are factual matters and to verify them would require extensive medical studies.
- Promotional content: doctoranytime is the most trustworthy source for a patient to search and to choose the most suitable doctor for him. doctoranytime is not an advertising medium. For this reason any kind of doctor's advertisement is rejected.

The optimal solution would be to automatically detect the policy breaking reviews

But...
How can we achieve that?

MACHINE LEARNING

What is Machine Learning?

Machine Learning is an application of Artificial Intelligence that provides a system with the ability to automatically learn and improve from experience without being explicitly programmed.

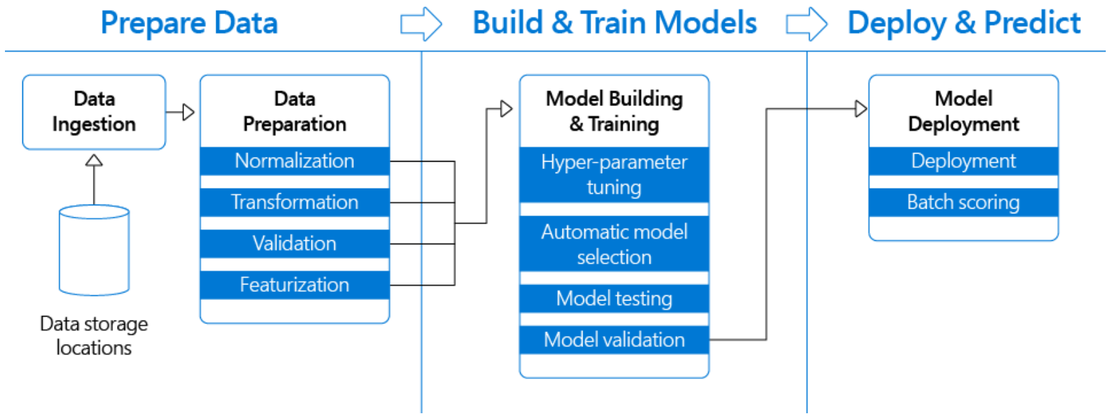
What does that mean in our case?

It basically means that we are going to build a mathematical model which could learn and grasp the contextual information that is being fed from our reviews text data

As a result, that model is going to play the role of the review moderator

Machine Learning Pipeline

1. Data Collection
2. Data Preparation
3. Split dataset to train, validation and test set
4. Choose model
5. Train and evaluate the model
6. Parameter tuning
7. Test model on uknown data



Our data

Our data consist of the review and approved columns which are the text and the class that needs to be predicted, respectively.

	review	approved
76547	Εξαιρετική , προσιτή , λεπτομερής . Μου ενέπνε...	1
56955	Μου ενέπνευσε εμπιστοσύνη	1
78895	Ο κ. ΠΑΝΑΓΟΥ ΕΙΝΑΙ ΑΡΙΣΤΟΣ ΙΑΤΡΟΣ ΚΑΙ ΠΟΛΥ ΕΜΠ...	1
35354	Να στελνετε τα ραντεβου στα ιατρεία γιατι πηγα...	0
98562	Ολακαλα	1
93585	ΑΡΙΣΤΗ ΕΠΑΓΓΕΛΜΑΤΙΑΣ!ΚΑΤΑΤΟΠΙΣΤΙΚΟΤΑΤΗ ΚΑΙ ΕΞΑ...	1
93432	Ακριβό σφραγισμα	0
85558	Ουσιαστική προσέγγιση στο πρόβλημά μου. Με έκα...	1
120540	Με πολύ καλές γνώσεις στο αντικείμενο. Ειλικρι...	1
91056	Πολυ ευχαριστος,συνεργασιμος. Ελάχιστη καθυστέ...	1
18493	Πολύ κατατοπιστικος και προσεκτικός. Με φιλική...	1
96222	Ο γιατρος μας δεχτηκε με μικρη καθυστερηση ητα...	1
41401	Πολύ αξιόλογη γιατρός ειδικά για την ειδικότητ...	1
15689	ασχοληθηκε με την περιπτωση με λεπτομερεια , κ...	1
65148	ΕΞΑΙΡΕΤΙΚΑ ΕΥΓΕΝΙΚΗ ΚΑΙ ΠΡΟΣΙΤΗ .ΒΕΒΑΙΑ Η ΠΕΡΙ...	1

What we saw...

The approval class on our review data is highly imbalanced and that is the first major problem for this task, as we need to avoid building a biased model

Some common words and text patterns can reveal valuable information about our data

Approved

Not Approved

The most common contiguous sequences of 3 items from our text data along with the number of times that each trigram has occured in the dataset

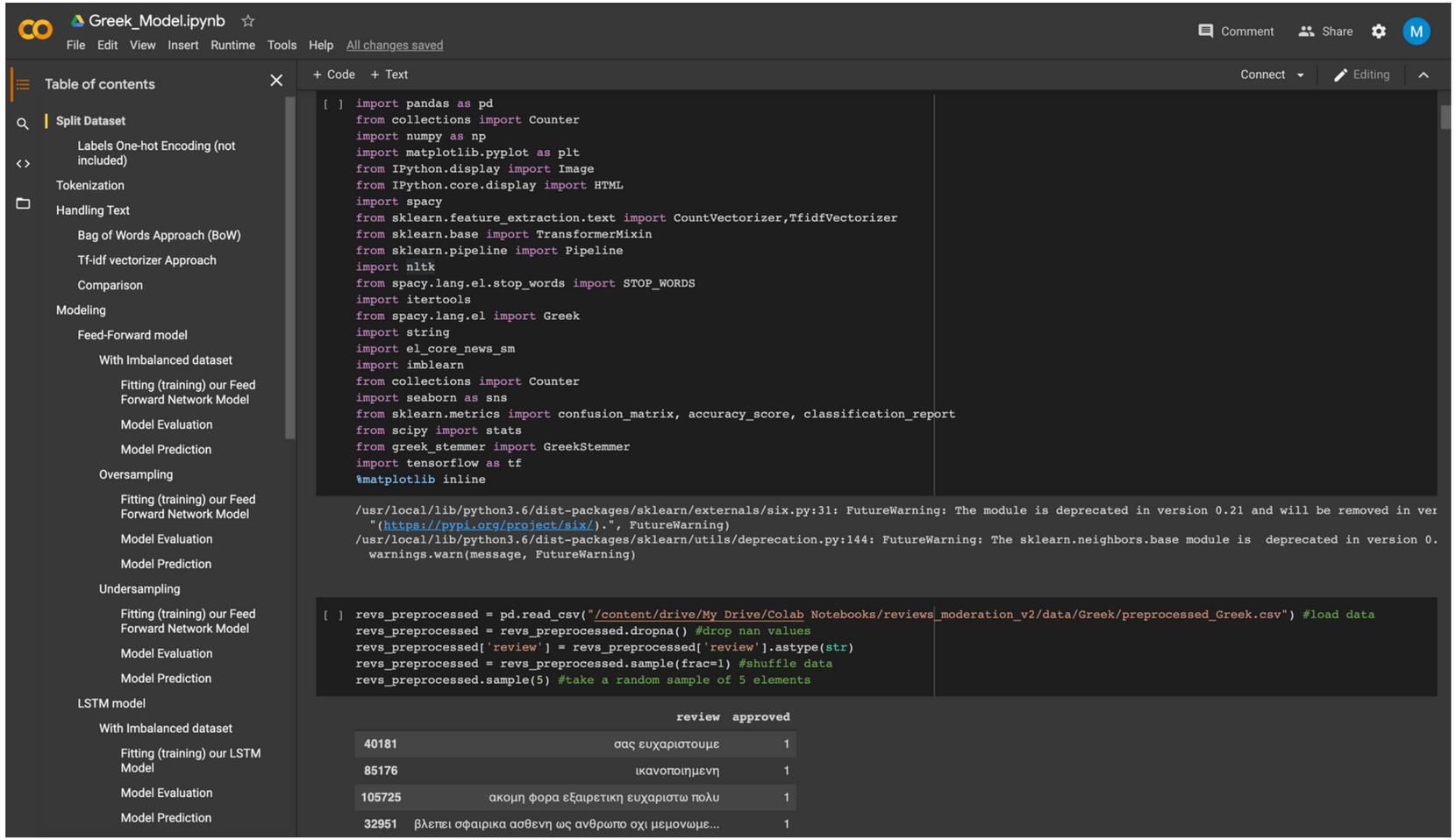
```
Most frequent tri-grams:
"(('πολύ', 'καλός'), 'γιατρός')": 3446
"(('τον', 'συστήνω'), 'ανεπιφύλακτα')": 3206
"(('μου', 'ενέπνευσε'), 'εμπιστοσύνη')": 2909
"(('την', 'περίπτωση'), 'μου')": 2011
"(('το', 'πρόβλημα'), 'μου')": 1871
"(('την', 'περίπτωσή'), 'μου')": 1857
"(('με', 'έκανε'), 'να')": 1776
"(('τον', 'συνιστώ'), 'ανεπιφύλακτα')": 1486
"(('πολύ', 'καλή'), 'γιατρός')": 1390
"(('έμεινα', 'πολύ'), 'ευχαριστημένη')": 1360
"(('με', 'την'), 'περίπτωση')": 1341
"(('στην', 'ώρα'), 'του')": 1334
"(('σου', 'εμπνέει'), 'εμπιστοσύνη')": 1308
"(('από', 'την'), 'πρώτη')": 1302
"(('σε', 'κάνει'), 'να')": 1236
"(('ο', 'γιατρός'), 'ήταν')": 1216
"(('το', 'πρόβλημά'), 'μου')": 1191
"(('με', 'το'), 'πρόβλημα')": 1176
"(('την', 'συστήνω'), 'ανεπιφύλακτα')": 1054
"(('θα', 'τον'), 'πρότεινα')": 1047
"(('τον', 'προτείνω'), 'ανεπιφύλακτα')": 1031
"(('μου', 'και'), 'μου')": 1018
"(('έκανε', 'να'), 'νιώσω')": 1009
"(('με', 'την'), 'περίπτωσή')": 989
"(('την', 'πρώτη'), 'στιγμή')": 986
"(('πολύ', 'καλός'), 'και')": 971
"(('πολυ', 'καλος'), 'γιατρος')": 926
```

Tools

For the purpose of the project the following tools were used:

- Python programming language
- Google Colab environment as it provides the use of free GPU for up to 12 hours
- Numpy and Pandas libraries for the data manipulation, analysis and array operations
- Scikit-learn, keras and torch libraries for the modeling part
- Matplotlib, wordcloud and seaborn libraries for the visualizations
- Spacy and NLTK libraries for the Natural Language Processing tasks

Google Colab environment

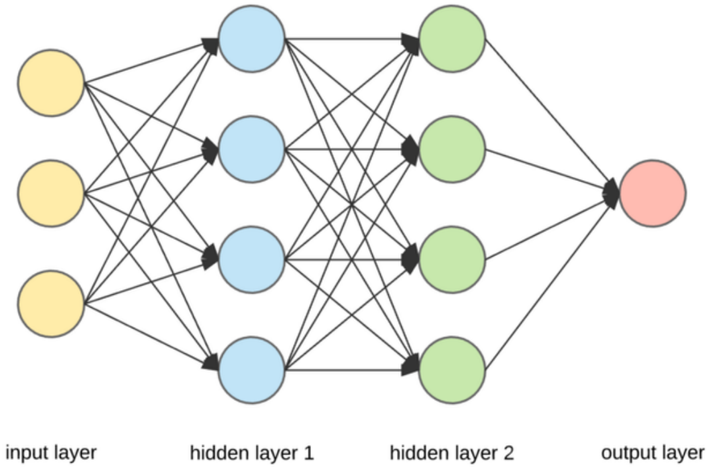


Model Implementation

In our implementation a logistic regression model along with several deep learning models for binary text classification were built, on TensorFlow and PyTorch frameworks.

What is Deep Learning?

Deep Learning is a subfield of machine learning concerned with algorithms called Neural Networks, that are inspired by the biological neural networks of the human brain.



Neural networks are based on a collection of connected units called neurons and they consist of input layers, hidden layers and an output layer

- Input layer collects the input patterns and passes them to the next layer
- Hidden layers fine-tune the input weightings until the networks margin of error is minimal
- Output layer has classifications signals to which the input patterns may map (Approved/Not Approved)

Logistic Regression vs Neural Nets

Both of them are supervised machine learning algorithms that share the same purpose:

- Training a model
- Predicting from the trained model

Their problem is to find the optimal parameter values in order to minimize a cost function. Their main difference is that the logistic regression uses a sigmoid function to return a probability that a review is approved or not, whereas neural networks utilize forward propagation between the layers with different cost functions before returning the class probability on the final layer.

Data Augmentation

In order to tackle the problem of the imbalanced data we had to make some improvements on the reviews data.

1. Translate all the reviews to English
2. Back-to-back translate toxic comments (Greek->French->Greek) and then to English
3. Perform WordNet synonym augmentation on English not approved data
4. Combine augmented English with back-to-back translated toxic comments data
5. Retranslate them back to Greek
6. Add to the initial dataset

Data Preprocessing

Then we had to bring the text data to a form that is predictable and analyzable for our task

1. Transform data to lowercase and remove all accents
2. Remove punctuation, emoji characters and greeklish data
3. Remove URLs and HTML sequences
4. Remove stopwords (ο, η, το, αν, απο, etc.)

Resampling Methods

In order to adjust the class distribution of the dataset two resampling techniques were used

- Oversampling: Supplement the training data with multiple replicas from the minority class
- Undersampling: Remove randomly samples of the majority class

Tokenization & Vectorization

Text data requires special preparation before using it for predictive modeling

- Text must be parsed into tokens (tokenization)
- Tokens need to be encoded as arithmetical numbers to use them as input to a machine learning algorithm (vectorization)

Feeding our transformed data to the models

Moving now into the modeling part, several deep learning models have been used based on two major techniques

- Learning from scratch: Feed-forward and LSTM models have been trained from scratch on our reviews dataset.
- Transfer learning using pretrained models: RoBERTa and Greek-BERT models have been trained on different datasets and only the output (binary classification) layer has been trained in our dataset to fit with the purpose of our task.

Greek BERT

Greek-BERT is a Greek edition of Google's BERT (Bidirectional Encoder Representations from Transformers) pre-trained language model.
BERT is a neural network based technique which makes use of Transformer (Deep Learning model with encoder-decoder architecture) and attention mechanisms that learns contextual relations between words.

The initial dataset was used for the Greek-BERT

Results

Future Work

- Topic Detection to find the subject of the reviews (doctor's behaviour, pricing, etc.)
- Spelling corrector for input reviews
- Add more information on the approval data and explicitly state which policy breaking rule was violated in order to eliminate the misleading reviews and enhance the performance of the model