# FinalReport

September 30, 2023

# 1 Sentiment Analysis on Machine Translated Icelandic corpus

Nemendur - Ólafur Aron Jóhannsson, Eysteinn Örn, Birkir Arndal

Leiðbeinendur - Hrafn Loftsson (hrafn@ru.is) - Stefán Ólafsson (stefanola@ru.is)

# 2 Contents

- 1 Sentiment Analysis on Machine Translated Icelandic corpus
- 2 Contents
- 2.1 Abstract
- 2.2 Introduction
- 2.3 Machine Translations
- 2.3.1 Google Translate
- 2.3.2 Miðeind
- 2.4 Pre-Processing and feature extraction
- 3 Baseline Classifier Evaluation
- 3.1 Support Vector Classifier
- 3.2 Logistic Regression
- 3.3 Naive Bayes
- 3.4 Testing
- 4 Next Steps
- 5 Burndown Chart
- 6 Risk Analysis 7 Status Meetings

#### 2.1 Abstract

Translating English text into low-resource languages and assessing sentiment is a subject that has received extensive research attention for numerous languages, yet Icelandic remains relatively unexplored in this context. We leverage a range of baseline classifiers and deep learning models to investigate whether sentiment can be effectively conveyed across languages, even when employing machine translation services such as Google Translate and Miðeind machine translation.

#### 2.2 Introduction

In this research endeavor, we utilized an IMDB dataset comprising 50,000 reviews, each categorized as either positive or negative in sentiment, with 25.000 being positive and 25.000 being negative. Our methodology involved the translation of these reviews using both Google Translate and Miðeind

Translate. Subsequently, we subjected all three datasets, including the original English version and the two translations, to analysis using three baseline classifiers. The primary objective was to investigate whether machine translation exerted any influence on the results of sentiment analysis and to determine the superior performer between Miðeind and Google translations. Our aim was to assess the transferability of sentiment across machine translation processes.

#### 2.3 Machine Translations

We employed the Google Translator API, which relies on Google's Neural Machine Translation featuring an LSTM architecture. Additionally, we utilized the Miðeind Vélþýðing API for the purpose of machine-translating the reviews. The Miðeind Vélþýðing API is constructed using the multilingual BART model, which was trained using the Fairseq sequence modeling toolkit within the PyTorch framework.

# 2.3.1 Google Translate

All the reviews were effectively translated using the API, and the only preprocessing step performed on the raw data was the removal of  $\langle br/ \rangle$ . The absence of errors during the translation process could be attributed to the API's maturity and extensive user adoption. Nevertheless, it's worth noting that the quality of Icelandic language reviews occasionally exhibited idiosyncrasies.

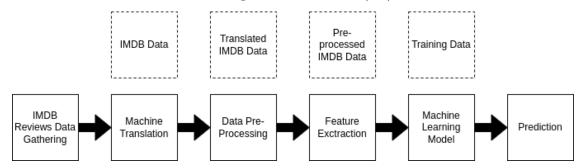
#### 2.3.2 Miðeind

The Miðeind Translator encountered challenges when translating the English corpus into Icelandic. To prepare the text for translation, several preprocessing steps were necessary. These steps included consolidating consecutive punctuation marks, eliminating all HTML tags, ensuring there was a whitespace character following punctuation marks, and removing asterisks. Subsequently, we divided the reviews into segments of 128 tokens, which were then processed in batches by the Miðeind translator.

## 2.4 Pre-Processing and feature extraction

The original English dataset was lowercased, tokenized and lemmatized, removed stop words and added a \_NEG prefix when a word was starting with negation to assist the vectorizer locating negative sentiments, the same method was applied on the Icelandic machine translated corpus as well.

Three baseline classifier pipelines were created that serve as a baseline metric for our scoring for English and machine translated Google and Miðeind datasets, all classifiers use TF-IDF vectorizer, which measure the frequency of a term in each document. It measure how important the term is across all documents. We see scoring of these terms in (3.2)



# 3 Baseline Classifier Evaluation

We utilized the classifiers available in the Scikit-learn Python package for implementing our machine learning models. These models were trained with their default parameters, and hyperparameter tuning was not conducted. It is important to note that superior results can be attained by fine-tuning the hyperparameters.

When assessing the statistical measures to gauge the model's performance, we applied equations 1, 2, 3, and 4.

$$Accuracy = \frac{TP + FN}{TP + FP + TN + FN} \tag{1}$$

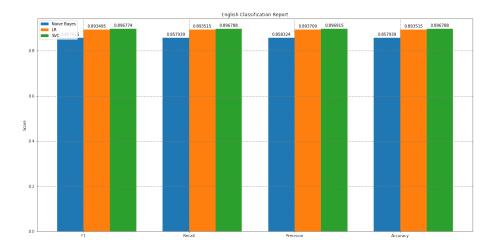
$$Recall = \frac{TP}{TP + FN} \tag{2}$$

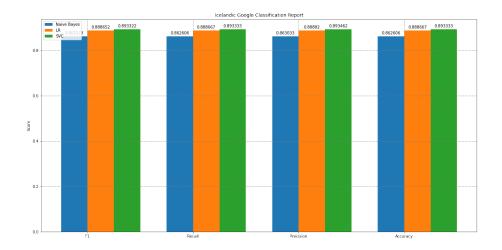
$$Precision = \frac{TP}{TP + FP} \tag{3}$$

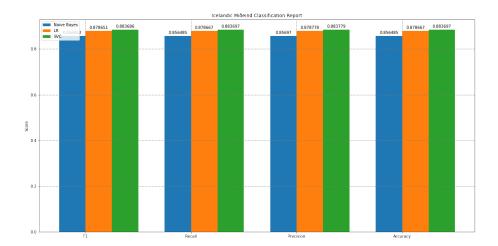
$$F1Score = \frac{2(Recall * Precision)}{Recall + Precision} \tag{4}$$

True Positive (TP) refers to correctly identified positive sentiments, while False Positive (FP) signifies incorrectly identified positive sentiments. True Negative (TN) denotes correctly identified negative sentiments, and False Negative (FN) represents incorrectly identified negative sentiments.

The data was divided into training and test sets, with 67% (33,500 reviews) allocated for training the models and 33% (16,500 reviews) reserved for testing the model's performance.







In this visual representation of the classification report encompassing all classifiers, we observe that Support Vector Classification (SVC) outperformed other models when applied to the data. All models were trained with 33,500 reviews and tested with 16,500. If we establish SVC as our baseline comparative model and employing a weighted F1 score as our evaluation metric, we can discern the following results across different datasets: In the English dataset, the F1 score reached 89.67%, the translated Miðeind dataset achieved an F1 score of 88.36%, and the Google dataset attained an F1 score of 89.33%. These figures suggest that sentiment analysis can carry across Machine Translation when utilizing state-of-the-art machine translation APIs. The loss in accuracy during translation is minimal, with only a 1.31% and 0.34% drop in accuracy, favoring Google's performance.

#### 3.1 Support Vector Classifier

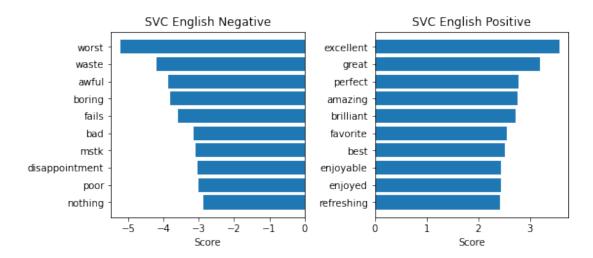
The SVC (Support Vector Classifier) was the best machine learning algorithm in classifying sentiment, it is a linear binary classification algorithm, where the result is defined as zero or one in binary models.

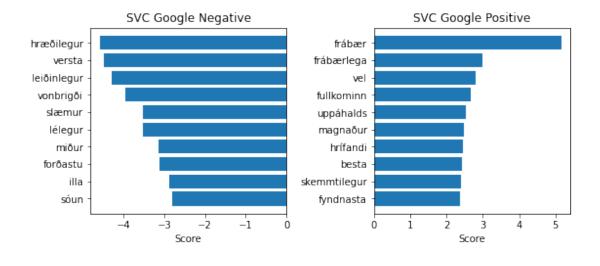
English Sentiment	Precision	Recall	F1-Score
negative positive	0.90	0.89	0.90
	0.89	0.91	0.90

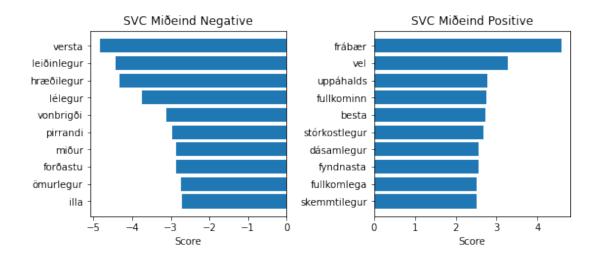
Google Sentiment	Precision	Recall	F1-Score
negative positive	0.90	0.88	0.89
	0.89	0.90	0.89

0.88 0.89

When we trained the class it gives us a list of coefficients that represent the relationship between the input variables and the output variable in the model. The coefficient can be interpreted as the relative importance of the word it's classified to, in this case negative or positive. In this chart we can see the top 10 negative and positive values, for a sentence to be positive in this case, it has to have a value of one.







# 3.2 Logistic Regression

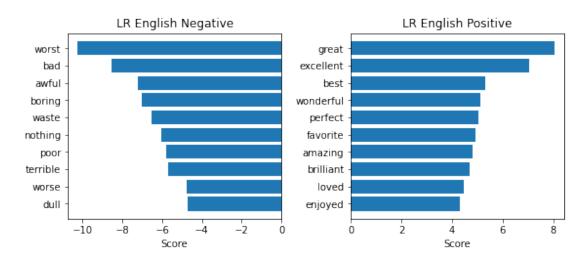
Logistic Regression is a binary classification algorithm, were the result is defined as zero or one in binary models.

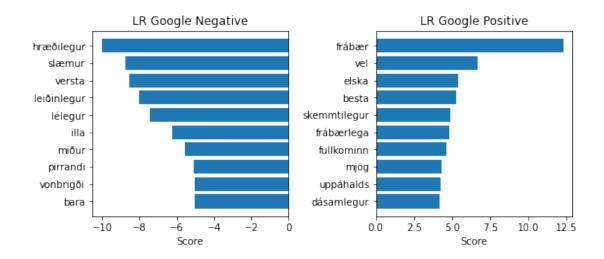
English Sentiment	Precision	Recall	F1-Score
negative	0.90	0.88	0.89
positive	0.89	0.91	0.90

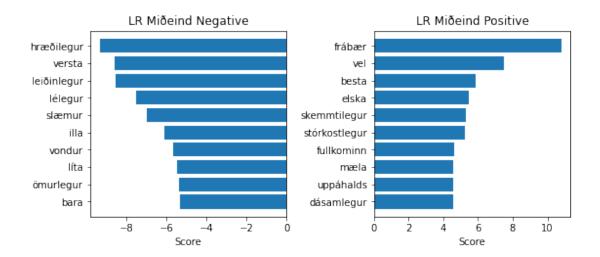
Google Sentiment	Precision	Recall	F1-Score
negative	0.90	0.88	0.89
positive	0.88	0.90	0.89

Miðeind Sentiment	Precision	Recall	F1-Score
negative	0.88	0.87	0.88
positive	0.87	0.89	0.88

When we trained the class it gives us a list of coefficients that represent the relationship between the input variables and the output variable in the model. The coefficient can be interpreted as the relative importance of the word it's classified to, in this case negative or positive. In this chart we can see the top 10 negative and positive values, for a sentence to be positive in this case, it has to have a value of one.







# 3.3 Naive Bayes

Naive Bayes is a classifier for multinomial models, although we employed it for binary classification, of all classifiers it performed the worst.

English Sentiment	Precision	Recall	F1-Score
negative	0.85	0.87	0.86
positive	0.87	0.84	0.86

Google Sentiment	Precision	Recall	F1-Score
negative	0.85	0.88	0.86
positive	0.88	0.85	0.86

Miðeind Sentiment	Precision	Recall	F1-Score
negative	0.84	0.87	0.86
positive	0.87	0.84	0.85

# 3.4 Testing

Additional testing on SVC was tried on Icelandic text from a website called https://www.officialstation.com/kvikmyndir, this is a website by an Icelandic author that seems to review movies in Icelandic(see screenshot)

\*\*\*\*\*

# Expend4bles (2023)

Ekki gott handrit, ekki góð samtöl. Ekki mikið verið að vanda sig. Stundum var augljóst að þetta var green screen eða tekið upp í stúdíói. Mætti jafnvel kalla þetta B-mynd. Var eiginlega fyndið hvað sum atriði voru kjánaleg/hallærisleg. En ágætis afþreying – með alveg því sem maður bjóst við: Slatti af action, slagsmálum, sprengingum og drápum.

#### Elemental (2023)

Skemmtileg fjölskyldumynd. Hugljúf og falleg. Virkilega flott – vel gerð. Vissi ekki að hún var frumsýnd á Cannes – lokaði hátíðinni og fékk 5-mínútna lófaklapp. #disneyplus

#### **The Founder (2016)**

Áhugaverð frumkvöðlasaga – en brútal hvað hann var óforskammaður til að ná sínu fram. Slatti af góðum leikurum.

#### Black Mass (2015)

Hellingur af flottum leikurum. Brútal mynd. Creepy týpur. Spennandi. Áhugaverð saga. Áhugavert að Jack Nicholson í The Departed var að hluta til byggður á Whitey Bulger.

### Those Who Wish Me Dead (2021)

Mjög spennandi mynd. Töff saga, fínasta plot. Eitthvað skemmtilega old school við söguna/handritið. Byggt á bók síðan 2014.

# Hidden Strike (2023)

Eitthvað off við handritið og samtölin. Gaman að sjá klassísk Jackie Chan bardagaatriði. Smá kjánalegt "grín action". #netflix

We tried random 10 reviews written by him to do more verification on the SVC classifiers, of the 10 reviews one review was marked incorrect (True Negative) by both classifiers, in addition to another review marked incorrect (True Negative) by Miðeind classifier, which makes the accuracy 90% for Google and 80% for Miðeind.

```
and present the property of Charley - Songies (law, geogle, predict([tx1][8])) - Noblesis ((ive, mideral predict([tx1][8])) - Test([tx1]))

For a proces input Tuniformality

For a procession in the process of the pr
```

# 4 Next Steps

Next steps are using classifiers that have scalable sentiment and start looking into deep learning models such as BERT, iceBERT and RoBERTa.

# 5 Burndown Chart

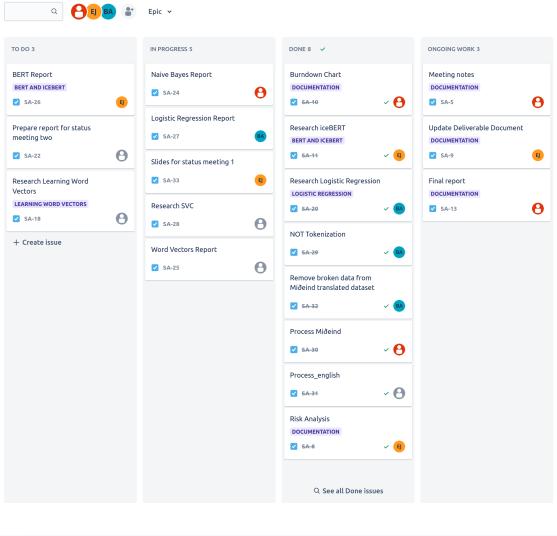
Given the research-oriented nature of our project, as opposed to corporate work, we opted for a Kanban approach rather than Scrum. We started well in advance, with initial preparations and research activities commencing in late July to early August. This timeframe allowed us to familiarize ourselves with the intricacies of machine learning, particularly since only one team member possessed prior experience in Machine Learning and Deep Learning.

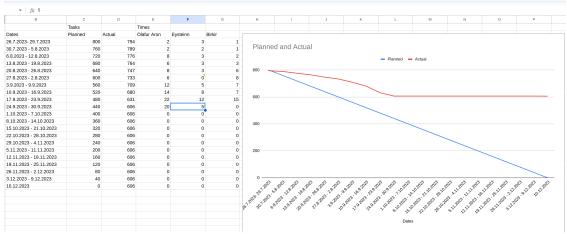
We've allocated a collective 40 hours per week for all team members, distributed across a span of 20 weeks, aiming to complete this project within this timeframe. This amounts to a total of 800 hours dedicated to the project. We expect the burndown to go under the planned line in October since we are picking up the pace but still keeping the 40 hours as a median.

The sum of the spent hours for each team member is - Ólafur Aron Jóhannsson (100) - Eysteinn Örn (44) - Birkir (50)

The reason Olafur Aron has accumulated more hours is due to the fact that he will be departing abroad in late November and returning in early December, during which period his hours will be reduced.

#### Main board





# 6 Risk Analysis

Risk	Likelihood (1-5)	Impact (1-5)	Responsibility	Mitigation Strategy
Resource Constraints (Time and Computing Power)	4	5	Eysteinn	Prioritize key features and models that are critical to the project. Consider using cloud computing resources.
Training stops/Computer crashes	4	4	Ólafur	Regular backups and distributed training could mitigate this risk.
Sprint/Project delay	3	5	Ólafur	Address the problem in the standup's and frequent reassessments.
Incompatibility of Translation APIs	3	3	Birkir	Have fallback methods for each API, and make the system modular to easily swap out one service for another.
Classifier Model Inefficiency	3	3	Eysteinn	Use baseline models for initial testing before using more complex models like BERT, roBERTa, and IceBERT.
Overfitting in Model Training	2	4	Birkir	Utilize techniques such as cross-validation and dropout layers.

Risk	Likelihood $(1-5)$	Impact (1-5)	Responsibility	Mitigation Strategy
Illness in team	2	4	Whole Team	Cross-training and comprehensive documentation can help other team members pick up the slack. Tries not getting other team members sick.
API Rate Limiting or Costs	2	3	Birkir	Caching translated data and batch processing could help in minimizing the number of API calls.
A team member quits	1	5	Whole Team	Having a documented and modular project architecture allows for easier transition of responsibilities.
External Dependency Failures (APIs down)	1	2	Whole Team	Have a contingency plan, such as a local translation model otherwise wait and focus on a different task

One risk has already happened and this is a team member quits, Júlíus decided to quit the project but because the Project Proposal was well defined we adjusted accordingly and the only impact is that our final produce may not be what was originally intended.

# 7 Project proposal

Included is a screenshot of the initial project proposal, it has not been mended in any way at time of writing, as it is still premature to establish that we have to amend the document.

# Heiti verkefnis: Viðhorfsgreining í vélþýddum íslenskum texta

## Tengiliður:

Hrafn Loftsson, hrafn@ru.is

#### Markmið verkefnis

Markmið verkefnisins er að þróa hugbúnað sem greinir viðhorf (e. sentiment) í íslenskum texta með beitingu gervigreindar/máltækni. Mismunandi gervigreindarlíkön verða þjálfuð og metin á íslenskum gögnum. Tvíundarviðhorf (e. binary) verða borin saman við viðhorfsgreiningu á skala (sem dæmi: 1-5). Í verkefninu verður jafnframt skoðað hvort munur sé á viðhorfsgreiningu (e. sentiment analysis) líkans sem hefur verið þjálfað á vélþýddum text (frá ensku yfir á íslensku) eða á texta sem hefur verið lagfærður eftir vélþýðingu eða jafnvel skrifaður frá grunni á íslensku.

### Ítarlegri lýsing

Í verkefninu verður leitast við að viðhorfsgreina íslenskan texta sem ekki hefur verið gert áður en er þekkt viðfangsefni erlendis. Til eru greinar þar sem líkön fyrir viðhorfsgreiningu hafa verið þjálfuð á gögnum eins og kvikmyndagagnrýni úr Internet Movie Database (IMDB). Þar sem engin opin íslensk gögn fyrir viðhorfsgreiningu eru aðgengileg verða gögn úr IMDB vélþýdd yfir á íslensku. Á sambærilegan hátt verður farið með gögn sem geyma viðhorfsgreiningu á skala, t.d. Amazon Product Reviews. Skoðaður verður munur á niðurstöðum mismunandi gervigreindarlíkana² sem þjálfuð hafa verið á vélþýddum textum og jafnframt borið saman við líkön sem þjálfuð hafa verið á upprunalegum enskum textum.

https://www.linkedin.com/pulse/sentiment-analysis-spanish-machine-translated-imdb-mondrag%C3%B3n-

diego/ <sup>2</sup> BERT, Learning Word Vector, Naive-Bayes

#### Afurð verkefnis

Í þessu verkefni verður til opinn íslenskur gagnagrunnur fyrir viðhorfsgreiningu og ýmis líkön sem þjálfuð verða á gögnunum. Skipta má verkefninu í nokkra hluta:

- Þjálfa og meta mismunandi gervigreindarlíkön eins og BERT, Learning Word Vectors, Logistic Regression<sup>5</sup>, Naive-Bayes<sup>6</sup> og Support vector machines.
  - Vélþýða enskan texta með Google Translate og vélþýðingu Miðeindar
  - Þjálfa og meta líkön á upprunalegum enskan texta með tvíundarviðhorfi
  - pjálfa og meta líkön á vélþýddum íslenskan texta með <u>tvíundarviðhorfi</u> Þjálfa og meta líkön á upprunalegum enskan texta með <u>skalanleg</u>u viðhorfi
  - Þjálfa og meta líkön á vélþýddum íslenskan texta með skalanlegu viðhorfi
  - Þjálfa og meta líkön á handvirkt lagfærðum texta með tvíundarviðhorfi
  - Þjálfa og meta líkön á handvirkt lagfærðaum texta meðskalanlegu viðhorfi
- Rannsaka hver munur er á frammistöðu mismunandi gervigreindarlíkanna sem þjálfuð eru á vélþýddum texta samanborið við frammistöðu líkanna sem þjálfuð eru á upprunalegum enskum textum.
- Rannsaka mun á tvíundarviðhorfi og skalanlegu viðhorfi
- Handvirkt þýða texta (eða lagfæra vélþýddan texta) á tilteknum fjölda gagnrýna frá IMDB og skoða hvort frammistaða líkanna breytist við meiri gæði texta
- Finna út hversu mikið af gögnum þarf til að þjálfa líkön til að ná "góðri" niðurstöðu

Mikið hefur verið skrifað um vélþýðingar og viðhorfsgreiningar með gervigreindarlíkönum á

Mikið hefur verið skrifað um vélþýðingar og viðhorfsgreiningar með gervigreindarlíkönum á tungumálum eins og hollensku, þýsku, albönsku, spænsku og fleiri tungumálum. Þetta hefur aldrei verið gert fyrir íslensku. Niðurstöður okkar verða gefnar út í skýrslu eða vísindagrein sem munu sýna hvort að vélþýðingar séu nógu góðar til að meta viðhorf, eða hvort að íslenska þurfi að vera hreinskrifuð til að meta viðhorf. Kannaðar verða niðurstöður mismunandi líkana á upprunalegan texta og íslenskan texta.

# **Umfang verkefnis**

Áætlað er að verkefnið taki um 800-1000 klst, verkefni fyrir 4 nemenda hóp yfir eina

# Þróunarumhverfi og tækni

Jupyter Notebook Github

# Hópameðlimir

Ólafur Aron Jóhannsson olafuraj21@ru.is Eysteinn Örn eysteinni 19@ru.is Birkir Arndal birkirh20@ru.is Júlíus Grettir Margrétarson juliusm21@ru.is

<sup>3</sup> https://huggingface.co/mideind/IceBERT

<sup>4</sup> https://ai.stanford.edu/~ang/papers/acl11-WordVectorsSentimentAnalysis.pdf

https://www.dlsu.edu.ph/wp-content/uploads/pdf/research/journals/jciea/vol-7-1/4aliman.pdf https://aclanthology.org/S14-2026.pdf

# 8 Meeting Notes

In addition to all data gathered we also tried to keep meeting notes as far back as 26. July.

. นาเนนาชูงเจน เหาะ หางเเจเเจชูเจนนเชูน

# Fundur 26.7.2023

Nemendur stofna hóp og ræða sín á milli hugmyndir að lokaverkefni, Discord rás er stofnuð

#### Fundur 6.8.2023

Nemendur hittast og ræða sín á milli hugmyndir sem þeir vilja sjá í Lokaverkefni, Eysteinn hefur verið að vinna í verkefni tengt gervigreind og máltækni og voru allir nemendur sammála að við viljum vinna í verkefni tengt því, hugmyndir sem komu upp á fundi voru meðal annars(á Ensku:)

- Poetry (Help students in literature classes)
  - Analyze poetry (merkja stuðla og höfuðstafi)
  - Generate poetry in Icelandic
- Help with Icelandic vocabulary and sentences. AI can invent new sentences or questions in areas where students are struggling and can receive feedback from the system
- GameQA extensions
  - Focus on more complicated questions
- Multiplayer language learning game
  - Game where users compete against each in Icelandic learning challenges
  - Can be gamified where you level up
  - Could even be tied to the Poetry and Icelandic vocabulary ideas (mark Stuðlar and Höfuðstafir/What is wrong/right with this sentence, etc)
- Sentiment analysis for Icelandic text (who is positive/who is negative, could be interesting
  to see what news stations is the most positive/negative)
- Icelandic text simplification tool
  - Could be a benefit to foreign, young and even native-Icelandic speakers to be able to have a text (usually written by a Lawyer) and parse it to something that is human readable (from lawyer-speak)

Farið var yfir landslagið í gervigreind og máltækni og hugmyndir ræddar. Ákveðið var að hittast aftur eftir þrjá daga og undirbúa tölvupóst til að senda á Hrafn Loftsson.

#### Fundur 3.8.2023

Nemendur ræða lítilllega á Discord og skipuleggja fund og ræða um lokaverkefnið

#### Fundur 4.8.2023

Fundur skipulagður

#### Fundur 6.8.2023

Miðeind, <u>huggingface</u> or fleiri auðlindir tengdar gervigreind eru skoðaðar<u>, iceBER</u>T og BERT módelið

#### Fundur 9.8.2023

Hittust nemendur og var Eysteinn með póst tilbúinn sem nemendur lásu yfir.. Var einnig rætt um Miðeind API og fleiri gervigreindar máltæknis hugmyndir sem koma til greinia, sendur var tölvupóstur á Hrafn með tillögum að verkefni, verkefnin sem voru valin voru eftirfarandi

- Text simplification, einföldun á flóknum íslenskum texta.
  - Einfalda og auðvelda tungumálið fyrir bæði útlendinga og Íslendinga (einfalda lagabálka og/eða lög).
- Gervigreind til að laga málfar og stafsetningu.
  - o Hins vegar er miðeind með svipað tól nú þegar sem kallast <u>yfirlestur</u>
- GameQA extension með einbeitingu á flóknum spurningum og svörum.

Kallliski et ilægt av laga pav svolla.

- Sentiment analysis Skoða hvort texti sé jákvæður eða neikvæður.
  - Gætum þá kannað hvaða frétta miðill er jákvæðastur/neikvæðastur.
- Ljóða sköpunar gervigreind Skapa og greina ljóð, finna stuðla, höfuðstafi og semja ljóð eftir prompt.
  - o getur hjálpað nemum í grunnskóla og <u>framhaldskóla</u> að skilja íslensk ljóð.
- Sjálfvirkur lestur, flokkun og áframsending á íslenskum tölvupósti.
  - það er að ef póstur er sent á hjalp@fyrirtæki.is og inniheldur setningu með orðinu "bókhaldskerfið" þá er það áframsent á bokhaldskerfi@fyrirtæki.is sjálfkrafa.

# Fundur 11.8.2023

Hrafn svarar pósti og leggur til að við munum taka að okkur viðhorfsgreiningu sem lokaverkefni, ræddu nemendur sín á milli að við værum sammála að þetta væri góður kandídat að verkefni.

## Fundur 14.8.2023

Ákveðið var að svara Hrafni og velja dagsetningu til að funda með honum

## **Fundur 16.8.2023**

Hrafn svarar og fundur næsta dag kl 10 var bókaður

# Fundur 17.8.2023 með Hrafn Loftssyni

Hrafn lýst best á sentiment analysis, munum við halda áfram með að gera lokaverkefnis-tillögu með það í huga

- Skoða aðra sem hafa þýtt frá ensku á annað tungumál
- Þýða imdb yfir á íslensku með vélþýðing frá miðeind og google translate
- Skali og/eða boolean á viðhorfsgreiningu
- Handvirkt þýða texta á 1000 review og skoða muninn hvort líkanið gefur aðra niðurstöðu fyrir
- handvirka breytingu (væri þá áhugavert að hafa skal í þessu tilviki)
- Logistic regression til að skoða sentiment analysis á móti tauganeti
- Íslenskur gagnagrunnur fyrir viðhorfsgreiningu
- Prófa mismunandi vélbýðandalíkön/gervigreindarlíkon
- Framlagið eru tilraunirnar sjálfar
- Hvaða aðferð notuðu þeir
- · Post-editing?
- Nota yfirferð frá Miðeind?
- Hversu mikið af gögnum þarf að þjálfa netið til að niðurstaðan sé góð?
   5000/10000/25000?
- Google-a Vélþýðingar á þekktum gögnum

# **Fundur 19.8.2023**

Byrjað er á Lokaverkefnis tillögu skjali fyrir verkefnið sem nemendur ákváðu er viðhorfsgreining á íslenskum texta

#### Fundur 21.8.2023

Farið er yfir skjalið og líka rætt mismunandi tauganet sem við ætlum að beita í verkefninu

# **Fundur 3.9.2023**

Júlíus og Ólafur Aron fóru yfir JIRA borð og ræddu um áhættugreiningu og skjölun á því, talað var um að halda um status á verkefninu á JIRA, ákveðið var að halda annan fund samdægurs því vantaði Birki og Eystein

# **Fundur 3.9.2023**

Allir nemendur fara yfir JIRA borð og búa til tösk sem þeir assigna á sig, við klárum að gera vélþýðingar og bætum við töskum sem við á, mælum okkur mót við Sigurjón um að spjalla á þriðjudag.

# Fundur 4.9.2023

Svar frá Sigurjóni að taka spjall næstkomandi dag

## Fundur 5.9.2023

Fundur með Sigurjóni, farið yfir stöðuna og sýnt JIRA borðið, rætt um scrum og kanban, verkefnatillaga kynnt og <u>útdrög</u> að önn sett upp.

# **Fundur 8.9.2023**

Rætt saman og skoðað hvar allir eru staddir, Naive Bayes, Word Vectors, BERT og Logistic Regression

#### **Fundur 9.9.2023**

Sigurjón ekki lengur umsjónarmaður í verkefni heldur Hrafn

# **Fundur 11.9.2023**

Stefán Ólafsson er líka leiðbeinandi í verkefninu, ákveðið að halda fund kl 9 næsta dag

# **Fundur 12.9.2023**

Fundur haldinn kl 9, nemendur ræddu við Stefán og Hrafn um stöðu verkefnis og var ákveðið að halda fundi kl 9 á þriðjudögum, rætt var um að meta muninn á vélþýðingum google translate og miðeind og skoða baseline <u>classifiera</u> eins og naive bayes, logistic <u>regreesion</u>

# **Fundur 16.9.2023**

Júlíus og Ólafur ræða saman og kemst Júlíus að þeirri niðurstöðu að hann er að íhuga að segja sig úr verkefninu.

# Fundur 17.9.2023

Birkir og Ólafur ræða um morguninn og fara yfir logistic regression og naive bayes. Júlíus og Ólafur ræða aftur saman og er Júlíus alveg farinn úr verkefninu og eru við þá orðnir 3, Birkir,

### **Fundur 18.9.2023**

Stefán Óla fær stöðuuppfærslu sem er svohljóðandi Sælir @stefanola , staðan er þannig að ég og Birkir erum komnir á gott ról með þá classifiera sem við töluðum um, þá Naive Bayes(Multinomial) og Logistic Regression, við höfum forunnið Google Translate gagnasettið með því að lowercasea, lemma og fjarlægja stopporð og trainað módelin og erum við að fá umbb 86-90% accuracy á baseline classifiera. Við ætlum að byrja á skýrslu sem mun fara yfir þær niðurstöður sem komnar eru, Eysteinn er búinn að vera vinna í iceBERT og talaði hann um í gær að mögulega þyrfti hann að þjálfa það frá grunni. Næst á dagskrá er að forvinna miðeindarsettið og líka keyra módelin á ensku gagnasettin til að bera saman, eftir það færum við að skoða SVC/word2vec. Það má líka nefna að ég og Júlíus ræddum saman um helgina og af persónulegum ástæðum hefur hann ákveðið að segja sig úr verkefninu og við erum þá þrír. Við höldum áfram með það sem lagt var með í byrjun en munum þá aðlaga tillöguskýrsluna til að endurspegla stöðuna.

# **Fundur 19.9.2023**

Rætt við Hrafn og farið yfir stöðuna, hann bendir á að muna tímaskráningu og byrja á skýrslu.

### **Fundur 20.9.2023**

Fundur kl 20, nemendur hittast og fara yfir hvað komið er, Naive Bayes, Logistic Regression og iceBERT, skoða að þurfa að endurtranslate-a miðeind því gögnin eru skrítin og eru að birta stjörnur.

# Fundur 21.9.2023

Eysteinn og Ólafur Aron ræða aðeins stöðuna, fara yfir transformera og lokaskýrslus

## Fundur 22.9.2023

Birkir og Ólafur Aron ræða um þýðingar og datasettið

# Fundur 24.9.2023

Nemendur ræða um BERT model og hvað progress er, þurfum að ræða við kennara til að leysa nokkra hluti

# Fundur 25.9.2023

Nemendur fara aftur yfir stöðuna fyrir stöðufund næsta dag

#### Fundur 26.9.2023

Fundur með Hrafn og rætt um stöðufund 3 Október