

---

Alma Mater Studiorum - Università di Bologna

# **NLP Project - Presentation**

## **Human Values Detection Behind Arguments**

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING  
Artificial Intelligence

### **GROUP MEMBERS**

*Alessio Cocchieri*

*Riccardo Grassi*

*Alessio Bonacchi*

*Marco Guerra*

Academic year 2022-2023

FEBRUARY

---

---

# Introduction

---

# Task

## Multi-label text classification:

Given a textual argument and a human value category, classify whether or not the argument draws on that category.

**Arguments** are given as:

- **Conclusion**

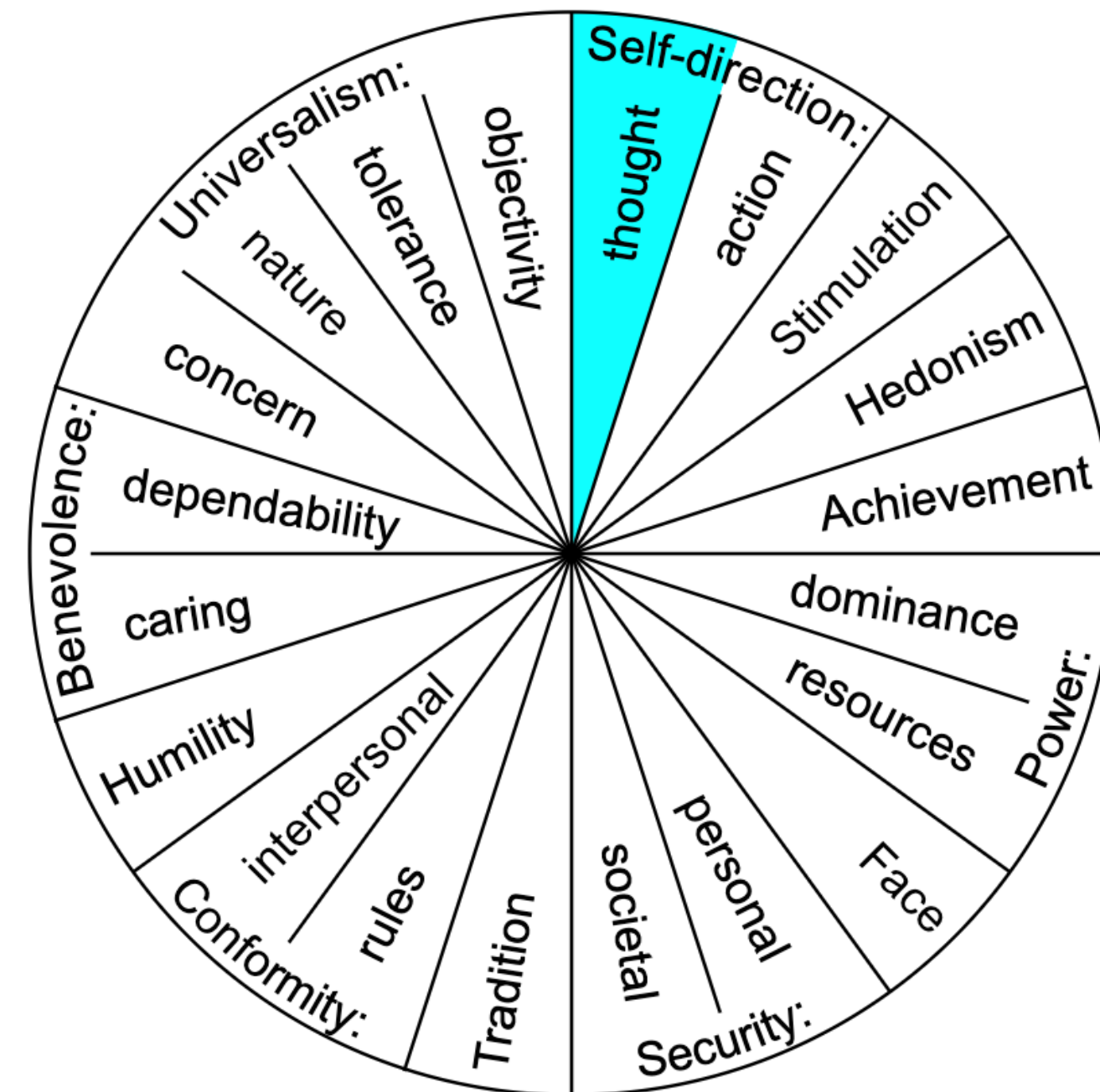
*We should prohibit school prayer*

- **Stance**

*againts*

- **Premise**

*It should be allowed if the student wants to pray as long as it is not interfering with his classes*



# Dataset

- We are using the data available on [Zenodo](#)
- [Human Value Detection 2023](#) is the competition which provide the source dataset

labels-train.csv

index	Argument ID	Self-direction: thought	Self-direction: action	Sti
0	A01002	0	0	
1	A01005	0	0	
2	A01006	0	0	
3	A01007	0	0	
4	A01008	0	0	

arguments-train.csv

index	Argument ID	Conclusion	Stance	Premise
0	A01002	We should ban human cloning	in favor of	we should ban human cloning as it will only cause huge issues when you have a bunch of the same humans running around all acting the same.
1	A01005	We should ban fast food	in favor of	fast food should be banned because it is really bad for your health and is costly.
2	A01006	We should end the use of economic sanctions	against	sometimes economic sanctions are the only thing that will get the corrupt governments to take action
3	A01007	We should abolish capital punishment	against	capital punishment is sometimes the only option to keep criminals from committing more crimes.
4	A01008	We should ban factory farming	against	factory farming allows for the production of cheap food, which is a necessity for families surviving on a low income.

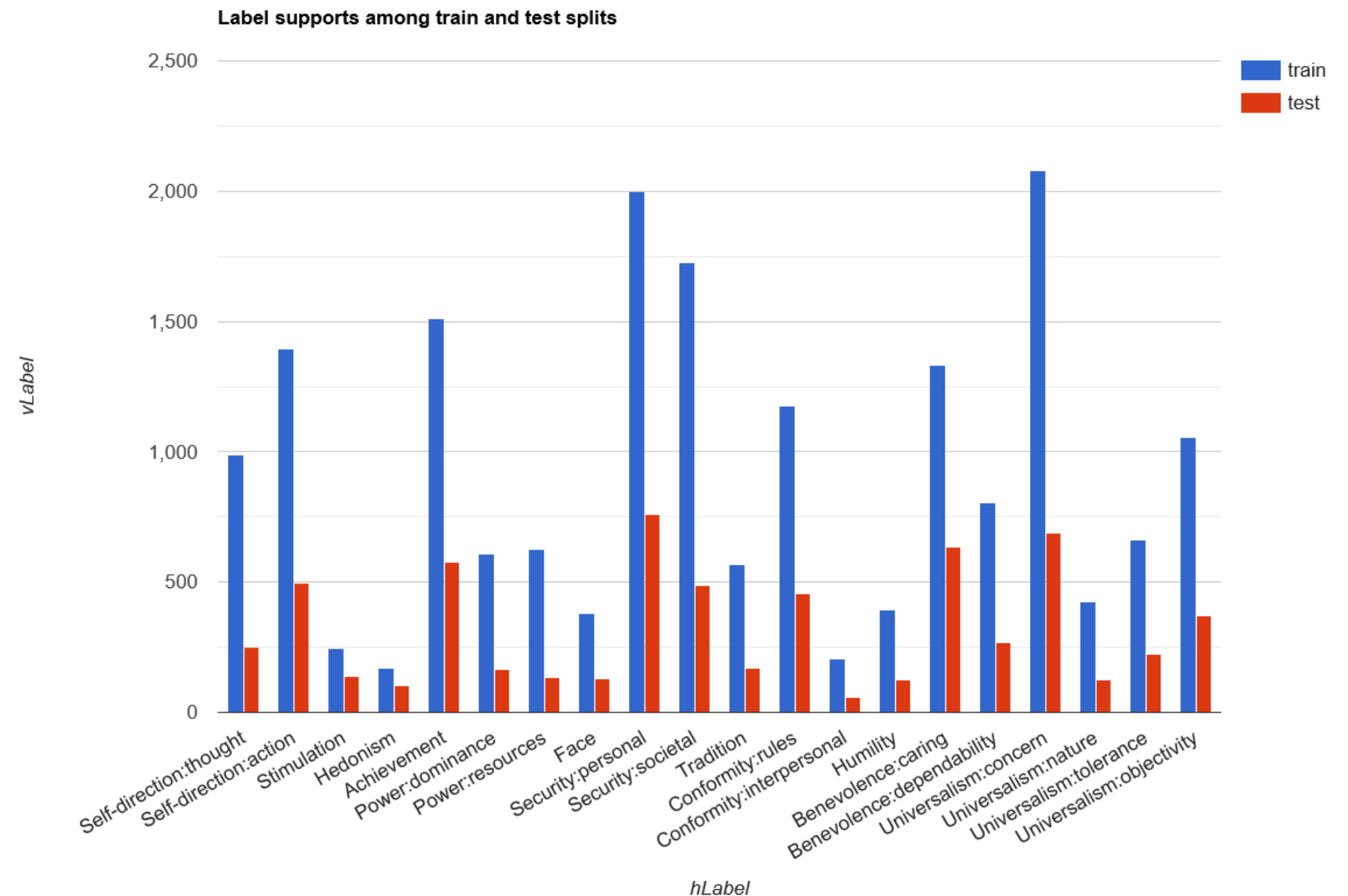
Since test data is provided without labels, we did not consider it for our analysis. In this regards **the performances of our models have been tested only on validation data.**

# Data overview

- **5393** data points in *train* split
- **1896** data points in *test* split

## High imbalanced class ratio

The number of instances for each label is not uniform, resulting in classes having a very high support and classes having a very low one.



# Selected Models

**SVM**

**BERT base**

**BERT large**

**DistilBERT**

**RoBERTa base**

**RoBERTa large**

**XLNet base**

**XLNet large**

---

# Support Vector Machine

---

# SVM

- Lowercase
- Stop words removal
- Punctuation removal
- Not-text removal
- Lemmatization

- TF-IDF  
{ max\_features = 5000 }

- Truncated SVD  
{ n\_components = 300 }

- OneVsRestClassifier
- SVC

**Input text: *Premises***

**Preprocessing**



**Vectorization**



**Dimensionality  
reduction**



**Classification**

## Parameters

- ▶ C = 18
- ▶ kernel = RBF
- ▶ gamma = 0.01
- ▶ class\_weight = balanced
- ▶ max\_iter = 10000



---

# Transformers

---

---

## Why BERT?

Try to outperform original paper results

## Why DistilBERT?

smaller, faster version of BERT that has been trained to have similar performance to the original BERT model

## Why RoBERTa?

It has been shown to perform better than BERT on a wide range of NLP tasks among which the current one.

## Why XLNet?

XLNet is a auto-regressive model which achieved state-of-the-art results on 18 tasks including text classification

# Just one preprocessing step...

Input text: Conclusion + Stance + Premise

index	text	list
0	We should ban human cloning in favor of we should ban human cloning as it will only cause huge issues when you have a bunch of the same humans running around all acting the same.	0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0
1	We should ban fast food in favor of fast food should be banned because it is really bad for your health and is costly.	0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0
2	We should end the use of economic sanctions against sometimes economic sanctions are the only thing that will get the corrupt governments to take action	0,0,0,0,0,1,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0
3	We should abolish capital punishment against capital punishment is sometimes the only option to keep criminals from committing more crimes.	0,0,0,0,0,0,0,0,0,0,1,0,1,0,0,0,0,1,0,0,0
4	We should ban factory farming against factory farming allows for the production of cheap food, which is a necessity for families surviving on a low income.	0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,1,0,1,0,0,0
5	We should fight for the abolition of nuclear weapons against nuclear weapons help keep the peace in uncertain times	0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,1,0,0,0

## ...NO more preprocessing needed

- Removal of punctuation, stop words and number
- Lemmatization



OVER  
PREPROCESSING



DEACRESE IN PERFORMANCE

---

**BERT**

**RoBERTa**

**DistilBERT**

---

# MODELS

**bert-base-uncased**

{ Batch size: 8, Epochs: 4 }

**bert-large-uncased**

{ Batch size: 8, Epochs: 4 }

**distilbert-base-uncased**

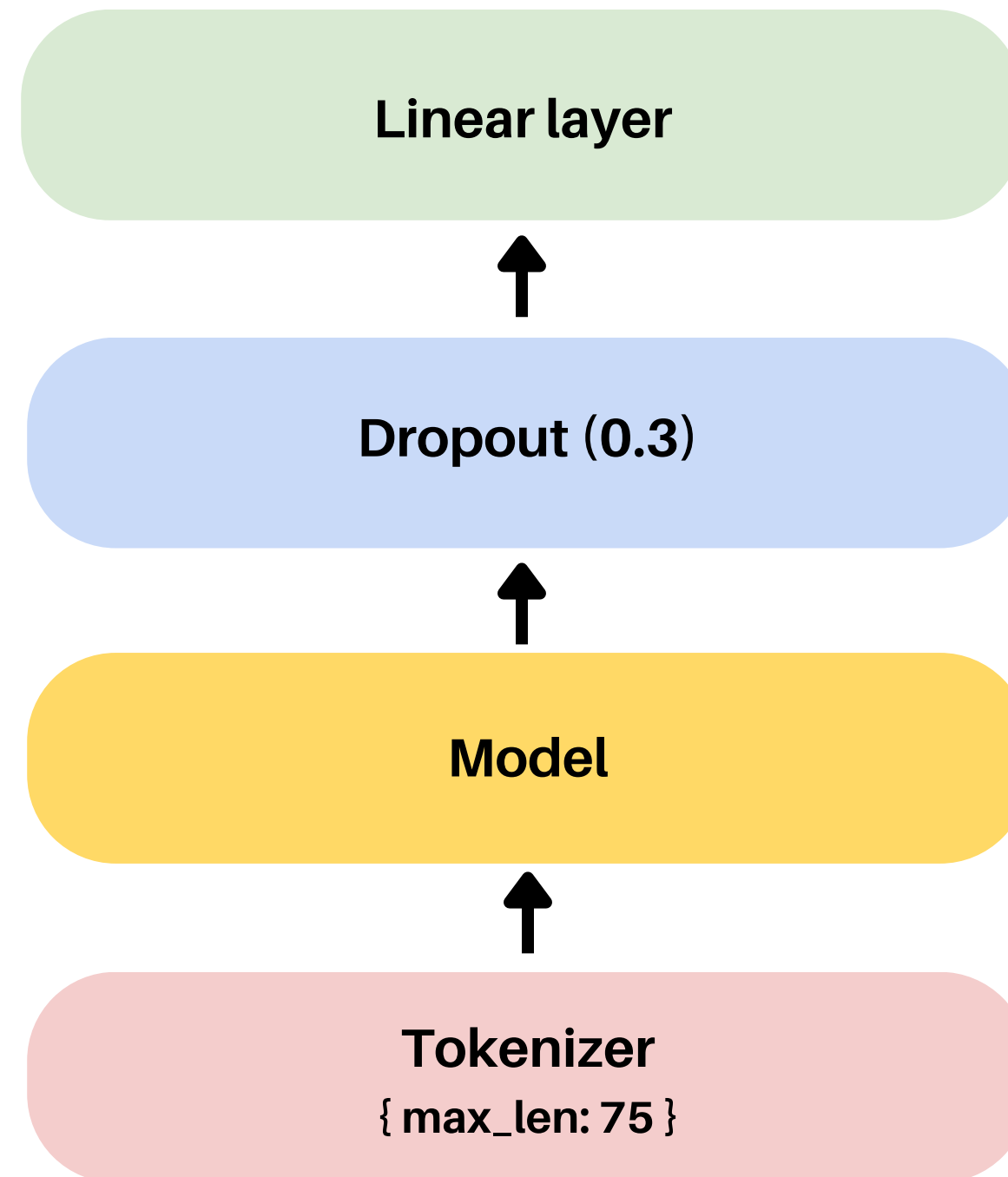
{ Batch size: 16, Epochs: 12 }

**roberta-base**

{ Batch size: 8, Epochs: 8 }

**roberta-large**

{ Batch size: 8, Epochs: 5 }



**loss**

BCEWithLogits

**optimizer**

Adam

**learning rate**

2e-05

■ NO linear scheduler  
■ with linear scheduler

---

# XLNet

---

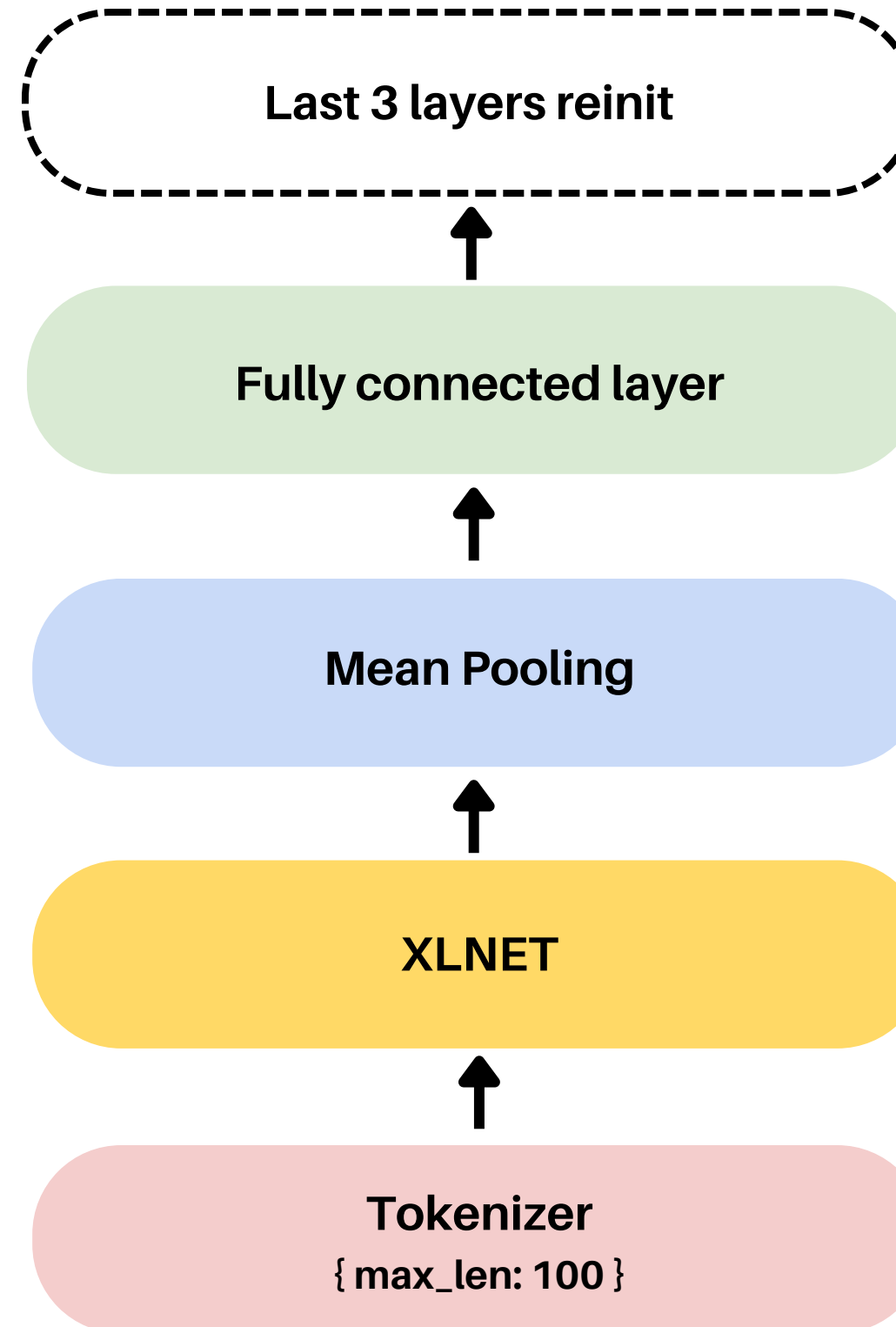
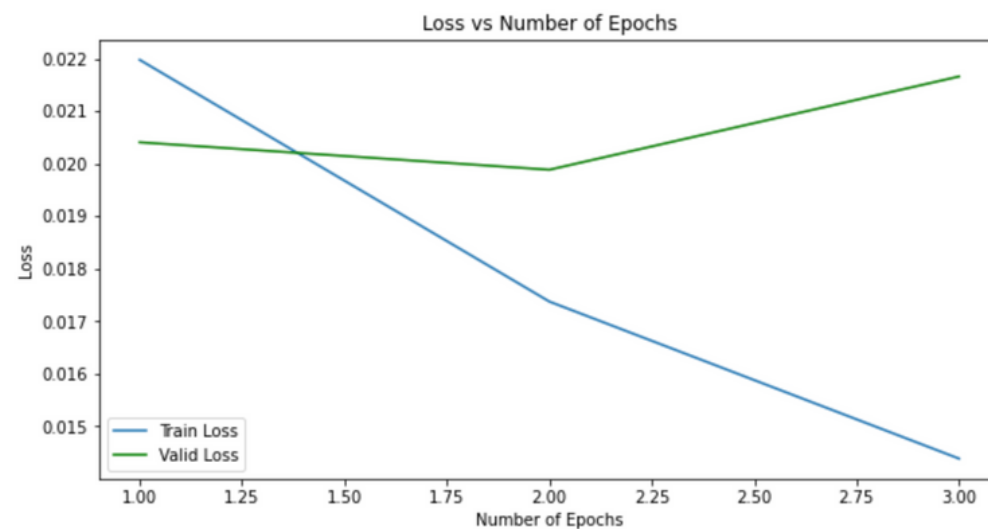
# XLNET

## xlnet-base-cased

{ Batch size: 16, Epochs: 3 }

## xlnet-large-cased

{ Batch size: 16, Epochs: 3 }



**loss**

BCEWithLogits

**optimizer**

AdamW

**learning rate**

2e-05

---

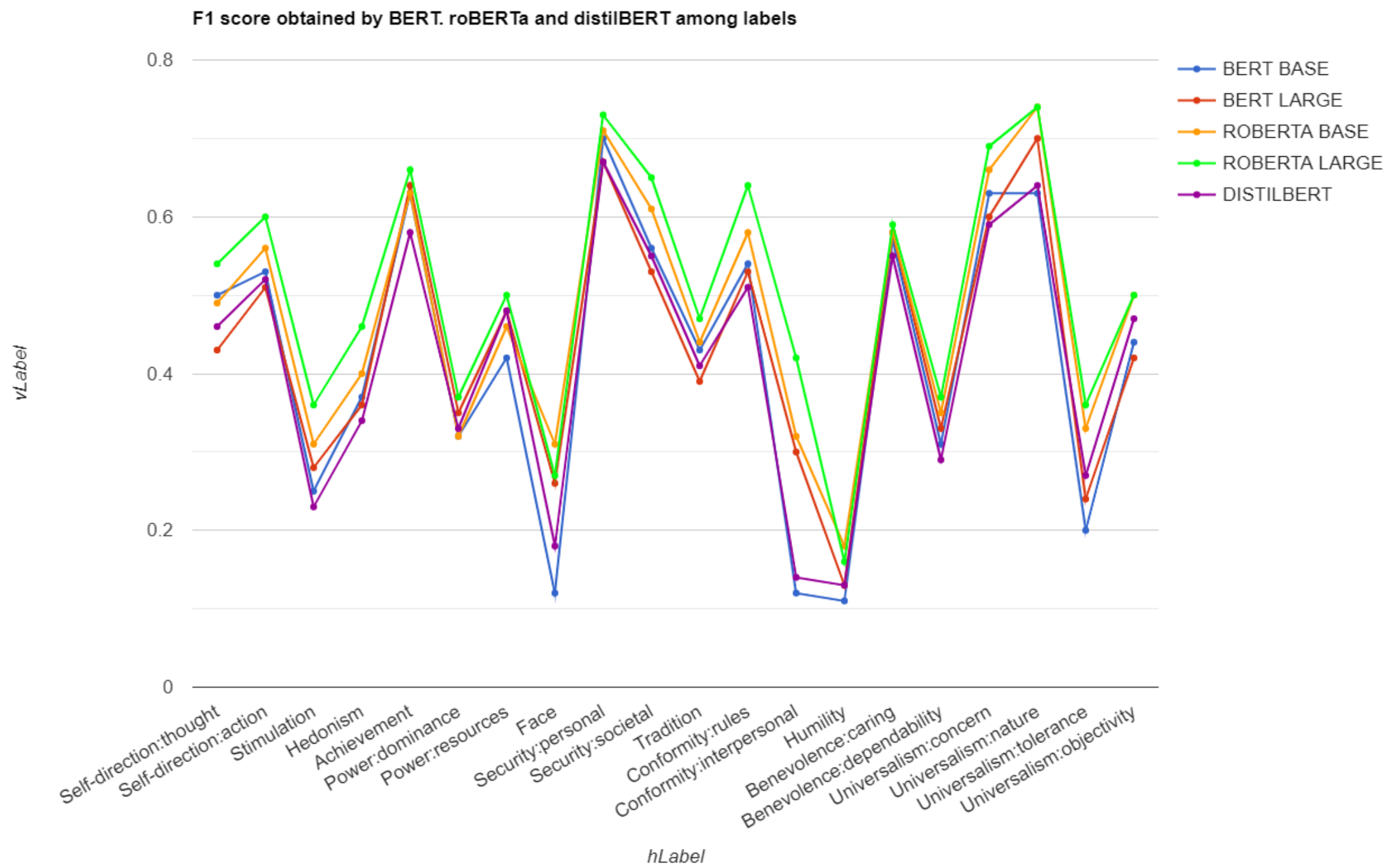
# Results

---



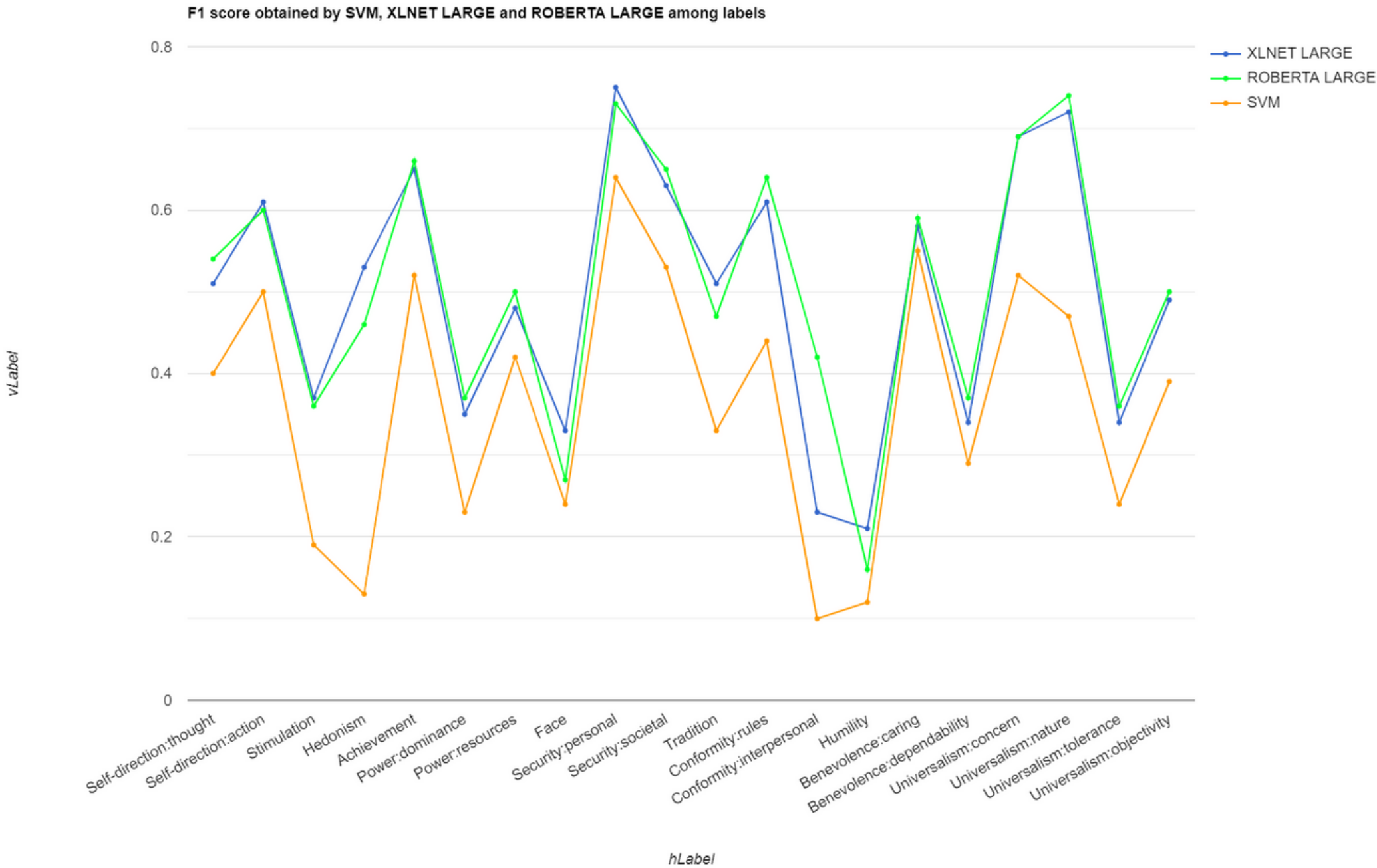
# BERT vs RoBERTa vs DistilBERT

MODEL	TEST F1	EPOCHS	BATCH SIZE
BERT base	0.42	4	8
BERT large	0.44	4	8
RoBERTa base	0.47	8	8
RoBERTa large	0.50	5	8
DistilBERT	0.43	12	16



# SVM vs RoBERTa-large vs XLNet-large

MODEL	TEST F1	iter/epoch	BATCH SIZE
SVM	0.37	10000	None
RoBERTa large	0.50	5	8
XLNet large	0.50	3	16



# Improvements respect to the literature

MODEL	PAPER (SVM)	SVM	Increase
F1	0.30	0.37	+23,3%

Performance **improved by more 20%** for SVM model

**Up to 47%** of  
performance  
improvement  
for transformers

MODEL	PAPER (BERT base)	BERT base	BERT large	Distil BERT	roberta base	roberta large	XLNET base	XLNET large
F1	0.34	0.42	0.44	0.43	0.47	0.50	0.44	0.50
	Increase	+23,5%	+29,4%	+26,4%	+38,2%	+47%	+29,4%	+47%

---

# Error Analysis

---

# Error analysis (TF-IDF)

General performances depends on the support for each class:  
**the more the support the more the F1 score, generally**

Not always true:

- **Universalism: tolerance**  
high support but low F1-score
- **Universalism: nature**  
low support but high F1 score

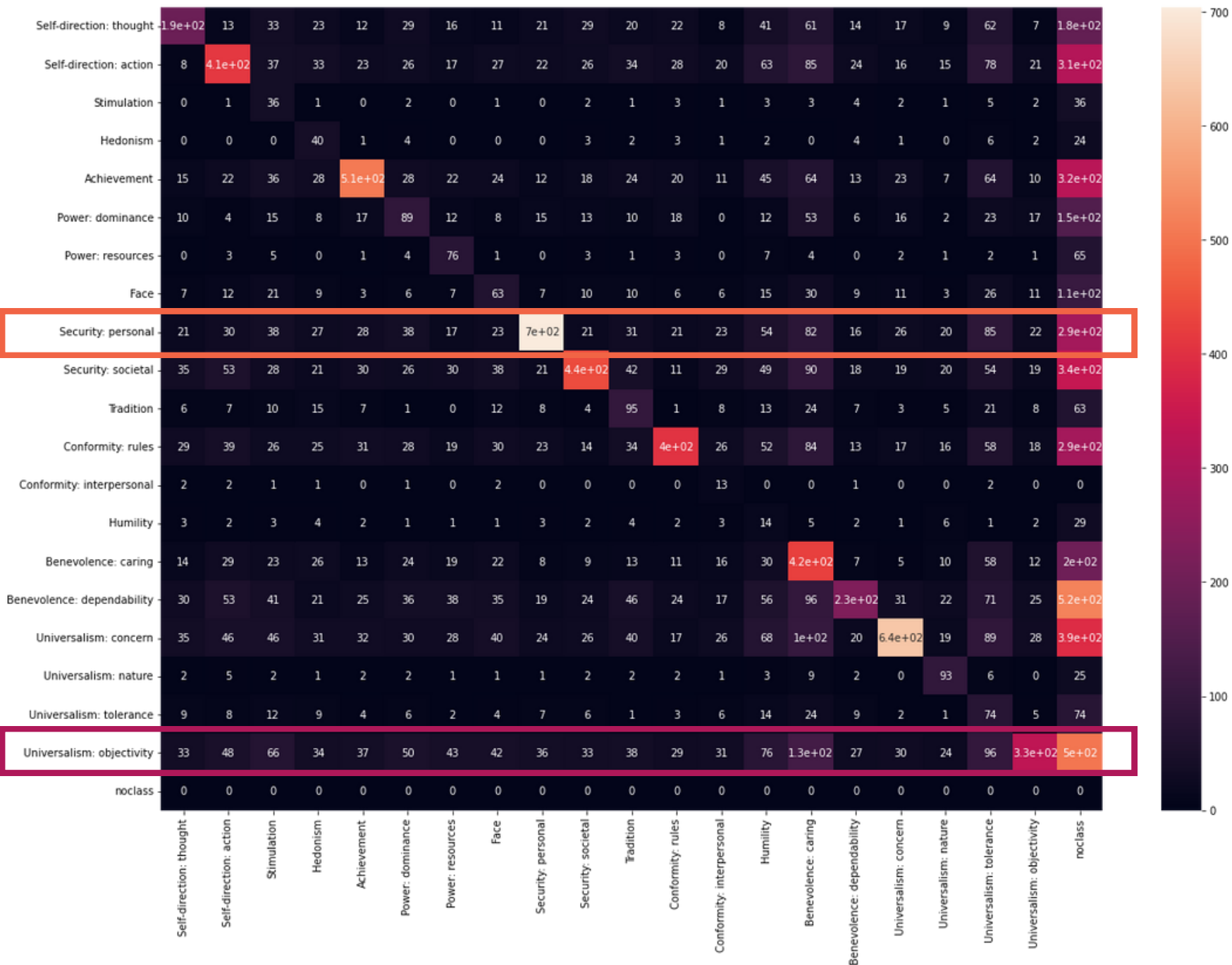
It can be explained **analyzing the TF-IDF scores**  
for the **15 most frequent words** of the selected class:

- **Universalism: nature**  
most frequent words show **high TF-IDF values only for this class**
- **Universalism: tolerance**  
many **other classes** show **high TF-IDF values**

ColumnName	natural	energy	nuclear	homeopathy	use	good	harm	people	promote	climate	farming	brings	make	change
Self-direction: thought	0.000000	0.000067	0.000000	0.000081	0.000000	0.000000	0.000000	0.000000	0.000043	0.000000	0.000000	0.000000	0.000000	0.000043
Self-direction: action	0.000042	0.000095	0.000000	0.000121	0.000000	0.000000	0.000000	0.000000	0.000040	0.000000	0.000000	0.000000	0.000000	0.000071
Stimulation	0.000051	0.000000	0.000000	0.000047	0.000000	0.000000	0.000000	0.000000	0.000037	0.000000	0.000000	0.000000	0.000000	0.000000
Hedonism	0.000000	0.000000	0.000000	0.000088	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Achievement	0.000011	0.000102	0.000026	0.000120	0.000000	0.000000	0.000000	0.000000	0.000049	0.000018	0.000000	0.000000	0.000000	0.000073
Power: dominance	0.000000	0.000000	0.000327	0.000107	0.000000	0.000000	0.000000	0.000000	0.000078	0.000000	0.000000	0.000000	0.000000	0.000104
Power: resources	0.000140	0.000533	0.000533	0.000011	0.000000	0.000000	0.000000	0.000000	0.000237	0.000000	0.000578	0.000000	0.000000	0.000068
Face	0.000000	0.000000	0.000000	0.000021	0.000000	0.000000	0.000000	0.000000	0.000034	0.000000	0.000000	0.000000	0.000000	0.000034
Security: personal	0.000051	0.000118	0.000118	0.000215	0.000000	0.000000	0.000000	0.000000	0.000069	0.000000	0.000107	0.000000	0.000000	0.000056
Security: societal	0.000000	0.000061	0.000273	0.000043	0.000000	0.000000	0.000000	0.000000	0.000039	0.000000	0.000000	0.000000	0.000000	0.000029
Tradition	0.000041	0.000000	0.000000	0.000075	0.000000	0.000000	0.000000	0.000000	0.000059	0.000000	0.000000	0.000000	0.000000	0.000000
Conformity: rules	0.000015	0.000035	0.000000	0.000059	0.000000	0.000000	0.000000	0.000000	0.000044	0.000024	0.000000	0.000000	0.000000	0.000066
Conformity: interpersonal	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Humility	0.000112	0.000000	0.000000	0.000026	0.000000	0.000000	0.000000	0.000000	0.000000	0.000179	0.000000	0.000000	0.000000	0.000122
Benevolence: caring	0.000053	0.000219	0.000073	0.000100	0.000000	0.000000	0.000000	0.000000	0.000054	0.000000	0.000099	0.000000	0.000000	0.000046
Benevolence: dependability	0.000000	0.000000	0.000111	0.000139	0.000000	0.000000	0.000000	0.000000	0.000018	0.000000	0.000000	0.000000	0.000000	0.000070
Universalism: concern	0.000000	0.000253	0.000210	0.000015	0.000000	0.000000	0.000000	0.000000	0.000033	0.000015	0.000000	0.000000	0.000000	0.000027
Universalism: nature	0.001176	0.002693	0.002609	0.000254	0.000000	0.000000	0.000000	0.000000	0.000536	0.001175	0.002170	0.000000	0.000000	0.000428
Universalism: tolerance	0.000054	0.000000	0.000000	0.000075	0.000000	0.000000	0.000000	0.000000	0.000099	0.000044	0.000000	0.000000	0.000000	0.000059
Universalism: objectivity	0.000000	0.000612	0.000382	0.000207	0.000000	0.000000	0.000000	0.000000	0.000097	0.000053	0.000000	0.000000	0.000000	0.000085



# Error analysis (MLCM)



## Multi Label Confusion matrix:

- shows the **distribution of false negatives (FNs)** from one class over other classes
- numbers on diagonals represent True Positives
- shows with which other classes the labels usually misclassified

## Examples

**Security:personal**

high F1 score, low number of FNs

**Universalism: objectivity** (roBERTa large)  
it is usually misclassified as Benevolence: caring, => **are we objectively caring?**

## Main issues

- Too small number of train samples
- Imbalanced classes
- Transformers might result as too complex architecture if not set up in the right way
- In particular, large models more prone to overfit

## Future improvements

- Models improvement might not be the main concern
- Increase the number of train samples
- Balance the ditribution of samples among each class

---

**Thanks for  
the attention!  
(is all you need)**

---