

Final Presentation – Classification of Text ML-Approach

Master Internship: Approaching Information System Challenges with Natural Language
Processing (IN2106, IN2130)

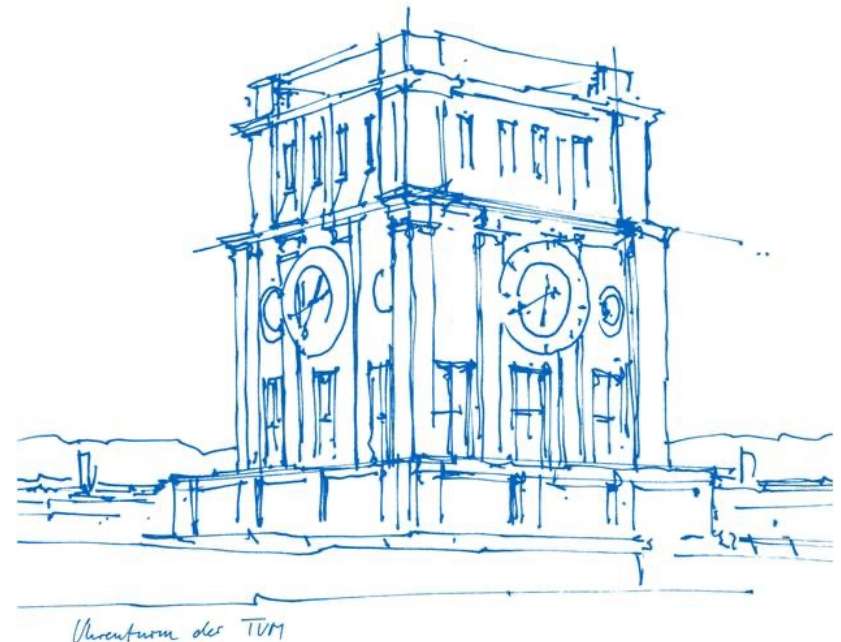
Patrick Ahrend

Garching, 30 January 2024



Agenda

- 1) Mission
- 2) How I planned the project
- 3) How it went
- 4) Explanatory UMAP
- 5) Results
- 6) Code Demo



Mission: Situation – Complication – Resolution



Situation

Companies have lots of business processes which need to be affirmative to legal requirements



Complication

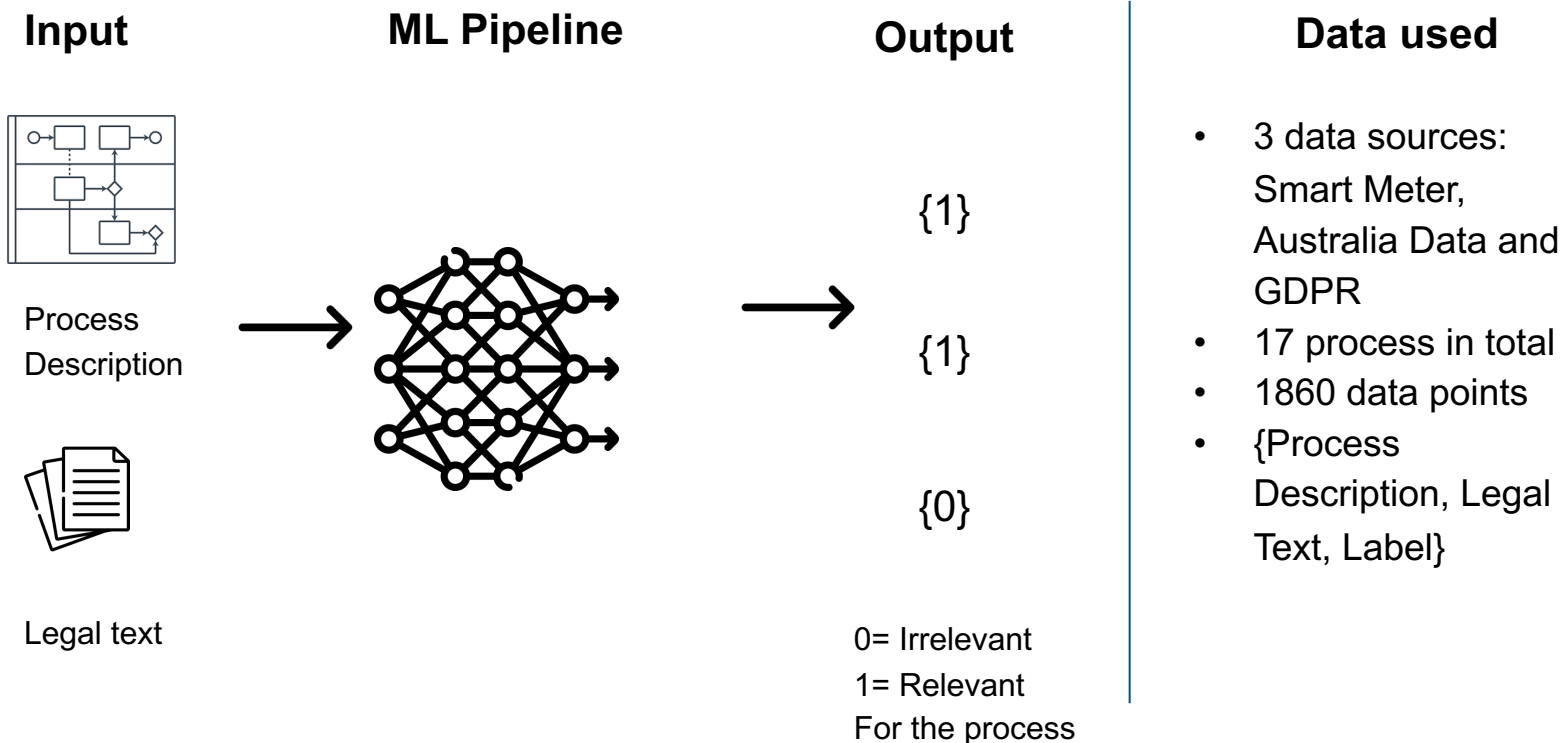
The process of legal work is quite labour-intensive and costly.



Solution

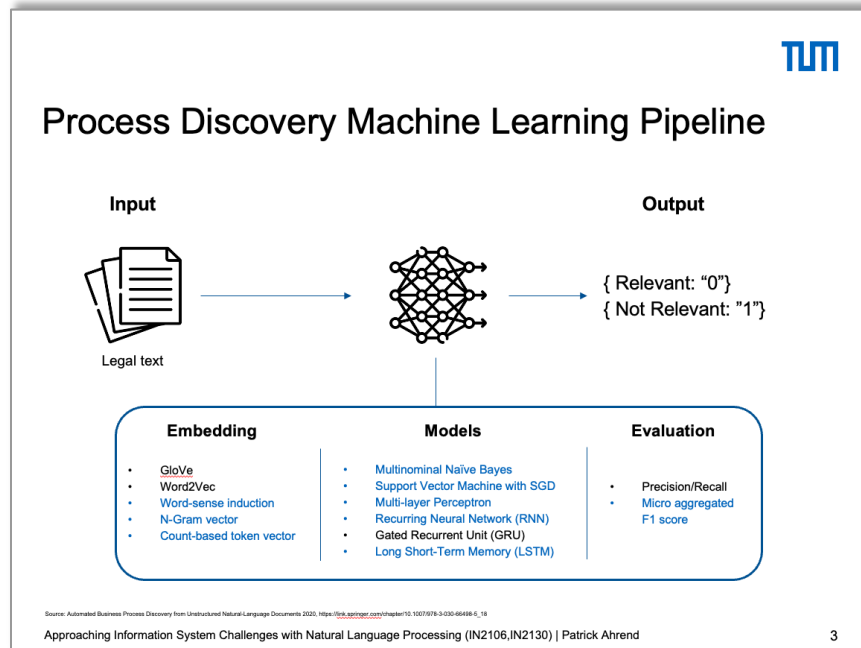
Semi Automation with ML model which is able to identify new regulations for human review

Solution: Machine Learning Pipeline for text classification

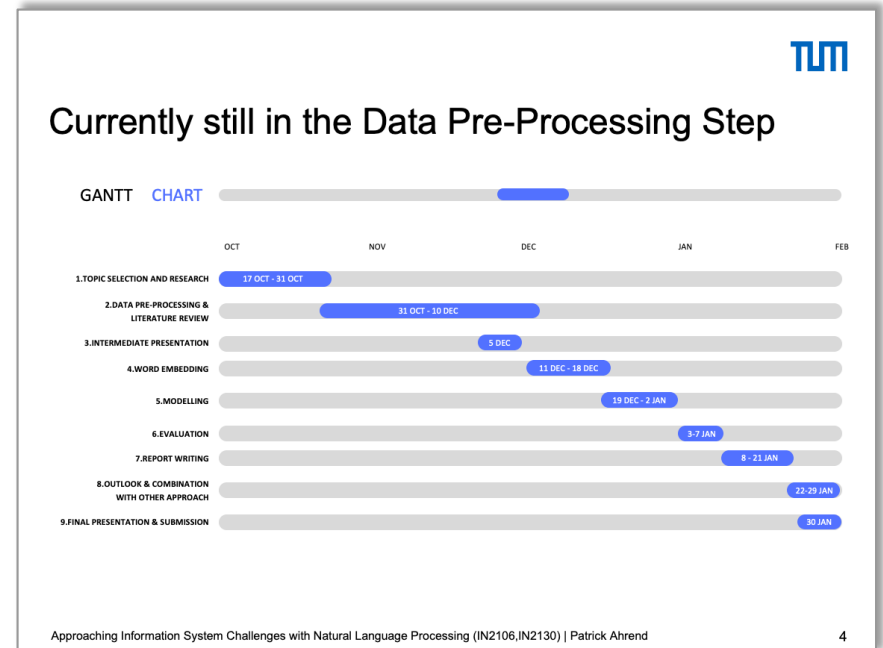


How the project was planned

Project Goal

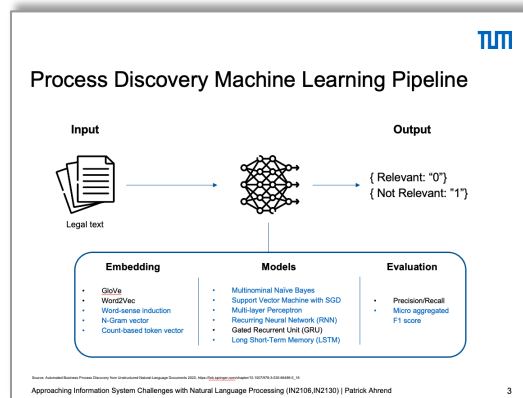


Project Timeline

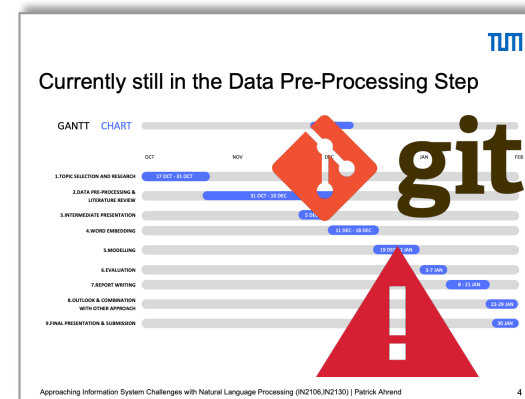
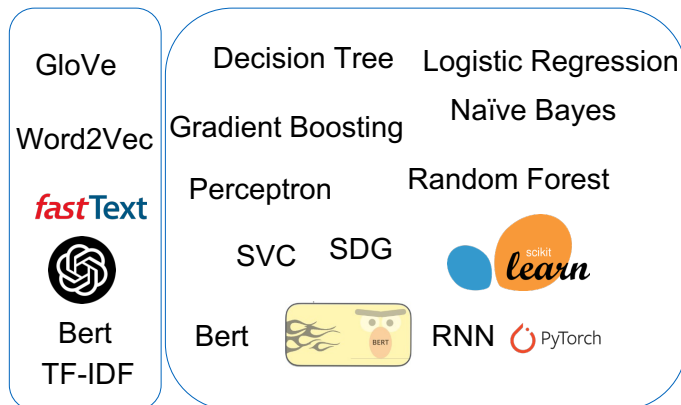
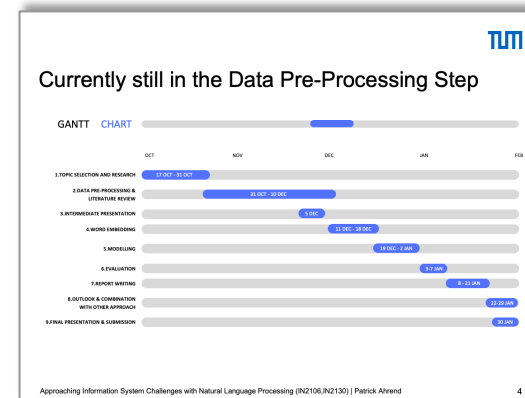


How the project actually happenend

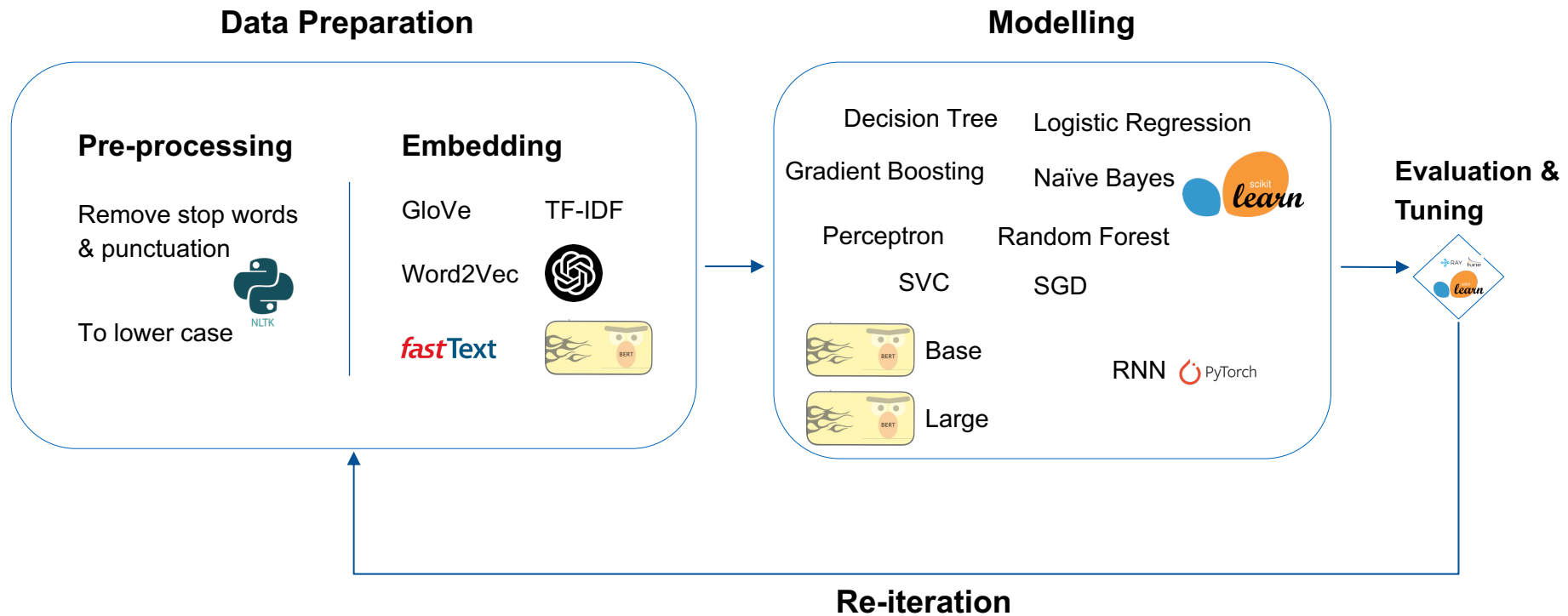
Project Goal



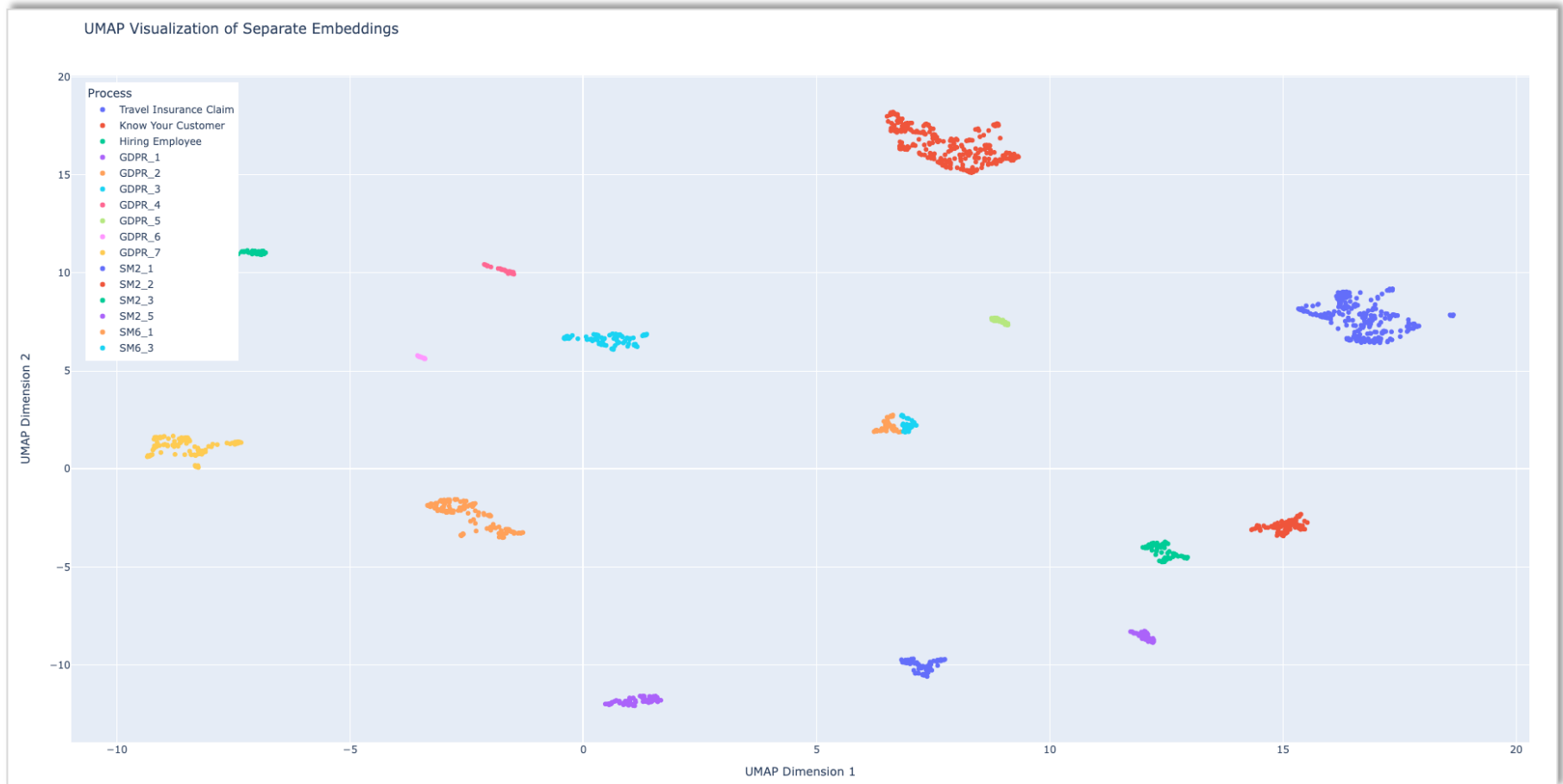
Project Timeline



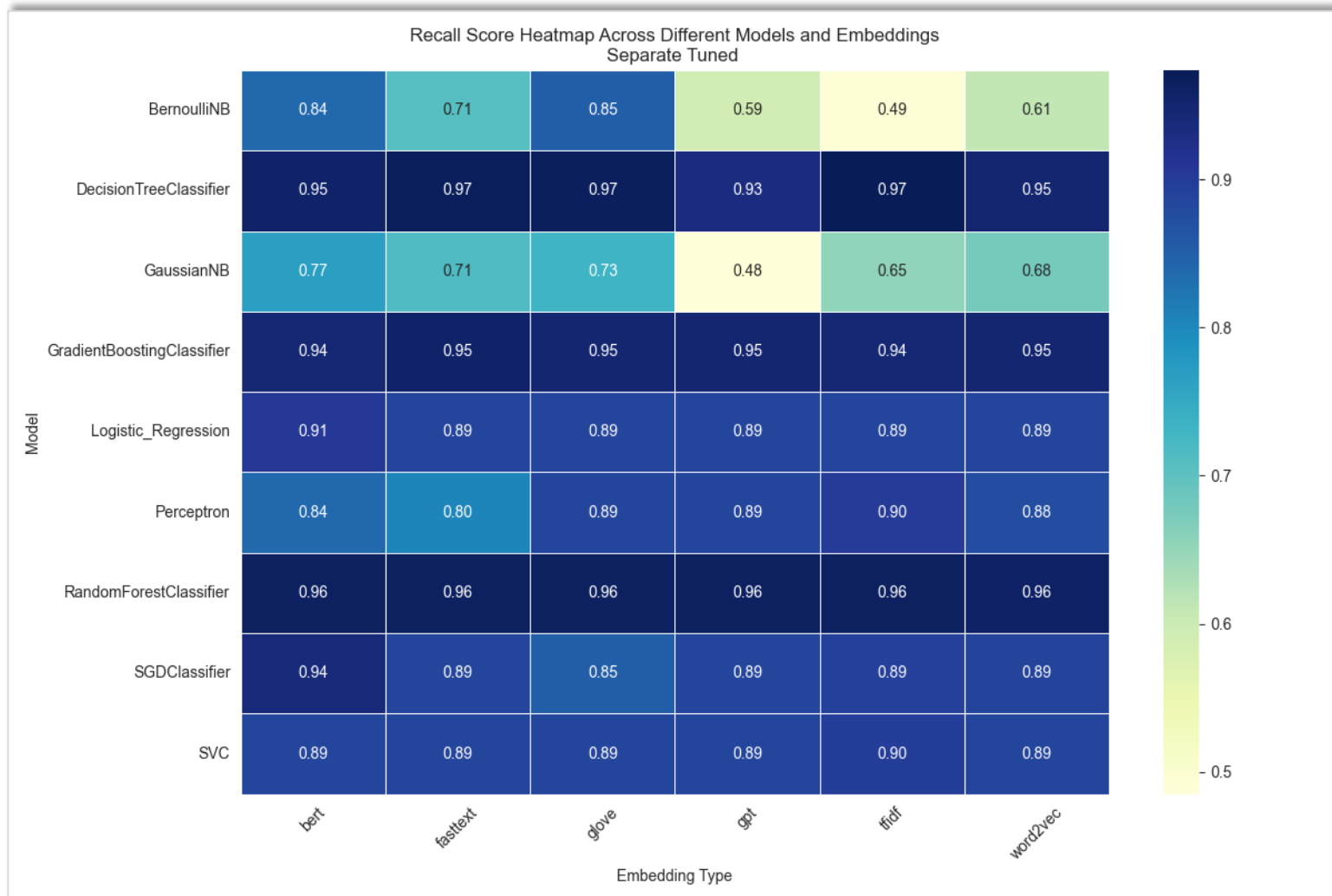
Machine Learning Pipeline for text classification



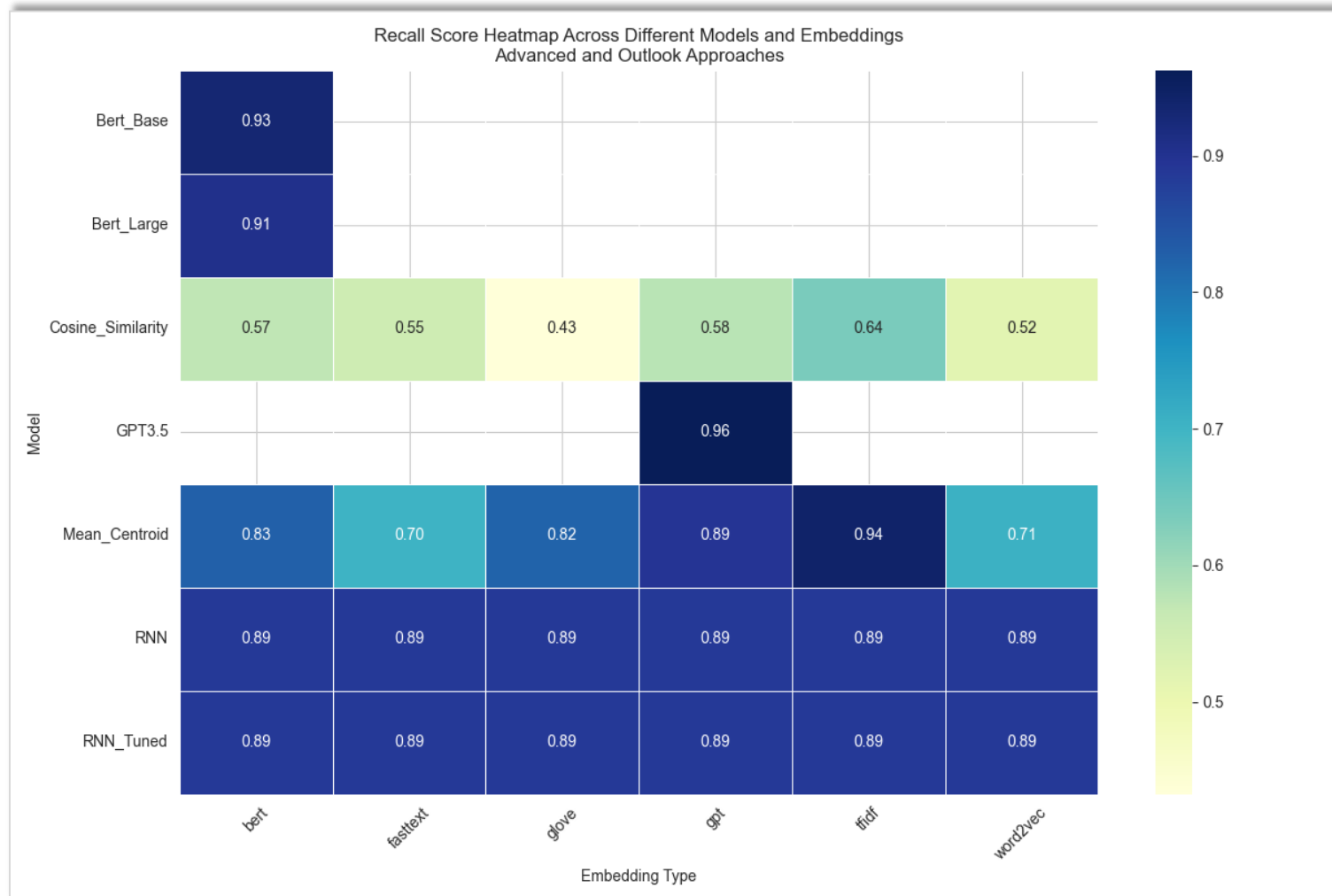
UMAP for Explanatory Data Analysis



Evaluation - Quantitatively I/II



Evaluation - Quantitatively II/II



Evaluation - Qualitatively

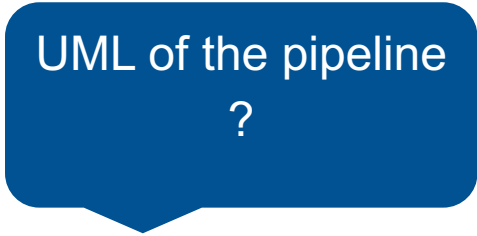
- 9 data points from travel insurance
- SGD, RNN, Perceptron, SVD failed to learn
- GPT and Bert predicted quite differently

Classification Results				Deploy
Text	Predicted	Actual	Status	
we our distributor and our service suppliers will be honest efficient fair transparent and timely in our dealing with you	0	1	Wrongly Predicted	
you may complain to we about any aspect of your relationship with we	0	1	Wrongly Predicted	
if you be entitle to claim an input tax credit in respect of a cost for which a claim be make or would be entitle to an input tax credit if you be to incur the relevant cost (i.e in replace a lost or steal item) the amount we would otherwise pay will be reduce by the amount of that input tax credit	0	1	Wrongly Predicted	
if any of the timeframe in this part be not practical due for example to the complex nature of your claim we will agree a reasonable alternative timetable with you if we can not reach an agreement on an alternative timetable we will provide detail of our complaints process	0	1	Wrongly Predicted	
we assumes you or your representative to tell us about your financial hardship so that we can work with you to discuss	0	1		

Code Demonstration

Thanks!
Questions/Feedback?

Implementation



UML of the pipeline
?

- Qualitative
- Quantitative