

Annotation Guidelines for Legal Judgment Prediction in Switzerland

Nina Baumgartner

1 Introduction

1.1 Annotation Goal

Recently [Niklaus, Chalkidis, and Stürmer \(2021\)](#) presented a diachronic multilingual (German, French, Italian) dataset for LEGAL JUDGMENT PREDICTION LJP including 85k Swiss Federal Supreme Court decisions. Using Hierarchical BERT, they achieved a Macro-F1 Score of up to 70%, considering penal law exclusively, they even achieved a score of up to 80%. But how accurately can legal experts predict the outcome of a court ruling by just using the facts section of a court case? Do they exceed the performance of the model in accuracy?

This annotation task has the goal to gather these legal expert judgment predictions. In addition, some free text explanations explaining the prediction will also be collected. These guidelines should help you the annotator to use the annotation tool and create consistent annotations. The guidelines you are currently reading are based on the work of [Reiter \(2020\)](#), [Leitner, Rehm, and Moreno-Schneider \(2019\)](#) and [Pustejovsky and Stubbs \(2012\)](#). They are a work-in-progress in collaboration with the annotators.

1.2 Dataset

The SJP dataset is split into training, validation, and testing set. For this annotation task, an almost balanced subset of the SJP containing 216 cases taken from the test and validation set was created. The dataset is deemed almost balanced because the 216 cases are equally distributed among the three languages contained in the Swiss judicial system German, French and Italian. Each language set contains twelve cases over six years (2015 until 2020). With each year having two cases per legal area¹: One with the verdict approved and one with the verdict dismissed. In addition, we tried to choose one case where the model was correct and one where it was not². Cases that were used in the judgment explainability task were discarded from this dataset. The reason is that the annotator may already know the judgment of these cases from the judgment explainability task.

1.3 Disclaimer

This document is a work-in-progress. If you have questions or find any errors in these instructions while doing the annotations please feel free to contact the maintainer. Please help with collecting examples to complete these guidelines.

2 Annotation Entities

This section describes the annotation entity you will be annotating by choosing a verdict and giving free text explanation. You will not have access to the entire court document. You should therefore base your decisions only on the text displayed to you by the annotation surface.

2.1 Fact Section

The text that you will be annotating is the fact section of a court ruling. The fact section contains the history of the case, legal claims, and the facts themselves. The facts form the foundation for the

¹The chosen legal areas are categorized as penal law, social law and civil law

²There are some instances of repetitions in the categories, especially in the Italian Dataset. This is why this dataset is only deemed almost balanced

considerations of the court. This section almost always begins with the word *'facts'* in the respective language of the court case (e.g. *'Sachverhalt'* in German) followed by several subsections which are marked with capital letters.

After reading the entire text you will annotate it by choosing the appropriate judgment of either **approved** or **dismissed** with a single-choice classification. Afterward you will explain why you chose accordingly. There might be cases where you have difficulties annotating. In such a case please reference the coming sections on annotation categories and choose the next most suitable category.

2.2 Explanations

Wiegrefe and Marasovic (2021) identify three types of explanations in the EXPLAINABLE NATURAL LANGUAGE PROCESSING (ExNLP) literature: highlights, free-text, and structured explanations. The explainability annotations for this task consist of structured and free-text explanations.

There will be a comment section on the annotation interface where you will be writing your explanation in the following structure. Please copy this template to the comment section and add your answers where \$ indicates it.

2.2.1 Comment Structure

\$Annotators name\$

Difficulty level for prediction: \$Number between 0-5\$

- Why was approval respectively dismissal chosen for this fact section: \$Your explanation\$
- Why was this difficulty level chosen: \$Your explanation\$

Add more questions to comment structure?

3 Annotation Categories

Niklaus et al. (2021) explained that originally the cases have been labeled with 6 labels: approval, partial approval, dismissal, partial dismissal, inadmissible, and write-off. To choose a judgment corresponding to a fact section you will have two possibilities. You can either approve a case or dismiss a case. There will be no option to deem a case 'inadmissible' or 'write off', the reason being that cases with this verdict were filtered out of SJP Corpus. Cases with a simultaneous ruling of approval and dismissal in the main request were also excluded from the dataset.³ So you will only read facts from cases where the main request has the same ruling for every sub-request.

In addition, you will be given several options for dealing with problematic cases, which should help to improve the dataset, these guidelines, and the annotations themselves.

3.1 Approval

This label is used when a fact section indicates that the request is deemed valid or partially valid respectively (i.e. the court "leans" positive). Please use approval and partial approval as synonyms and annotate a fact section indicating partial approval as approved.

3.2 Dismissal

This label is used when a fact section indicates that the request is denied or partially denied respectively (i.e. the court "leans" negative). Please use dismissal and partial dismissal as synonyms and annotate a fact section indicating partial dismissal as dismissed.

3.3 Problematic Cases

Problematic cases can occur. For now, we differentiate between two possible types of such cases.

³Niklaus et al. (2021) describe in their work that in practice, court decisions may have multiple requests (questions), where each can be judged individually.

3.3.1 Rejected Cases

If a case is badly tokenized⁴ or there is another formal error it should be rejected. Please state your reasoning in the comment window using the comment pattern below and reference the [Reject or Ignore a Case](#) section of this document for the details on how to properly reject a case.

3.3.2 Ignored Cases

If a case is for some reason unfit for the annotation it should be ignored. To ignore it please state your reasoning in the comment section and follow the steps explained in the [Reject or Ignore a Case](#) section of this document below.

3.3.3 Comment Structure for problematic cases

\$Annotators name\$

- Why was this case rejected respectively ignored: \$Your explanation\$

Add more questions to comment structure?

4 Implementation: How to Annotate the Dataset using Prodigy

This section explains how to use the annotation tool Prodigy⁵. We built a custom recipe for this task which lets you decide on the verdict of a court ruling given only the facts of it.

4.1 Access

The Prodigy instance can only be accessed via the University of Bern network. If you want to annotate from home you must use the VPN of the University of Bern⁶.

If you are connected to the university network you can access Prodigy via one of the URLs in the following three sections. Before you can start you will be asked to provide a *username* and a *password*, which will be given to you by the maintainer of the annotation process. After the login procedure, you should now see an overview of the case (see figure 1 below) and you can start with your annotation.

4.1.1 First cycle

The following links will be used for your pilot annotations (first iteration). If you completed the annotation on this dataset ignored and rejected cases will be replaced with other cases having the same legal area, year, and judgment. This process is ongoing until we reach 72 accepted cases per language. Note that you should insert your name at the end of the URL (instead of 'name').

- German case annotations: <http://fdn-sandbox3.inf.unibe.ch:14000/?session=name>
- French case annotations: <http://fdn-sandbox3.inf.unibe.ch:15000/?session=name>
- Italian case annotations: <http://fdn-sandbox3.inf.unibe.ch:16000/?session=name>

Note that sessions can be added dynamically by adding the suffix `/?session=name` to the url.

4.2 Annotate a fact section

Figure 1 displays the annotation surface on Prodigy. To choose the judgment for a facts section click on either 'Approval' or 'Dismissal' in the single choice box (see number [1] in figure 1). After choosing your judgment please comment on your decision in the comment section indicated with [2]. If you are happy with your annotation you can accept it by clicking on the green check labeled with [3] in Figure 1 and save it by pressing the save button in the left corner referenced by the number [4]. To see your progress you can look at the information displayed on the left (see the number [5] on Figure 1). Please do not forget to save your progress using the save button [3].

If you want to skip a case because you already annotated it. Please use the accept button [1] to get to the next case.

⁴Tokenized means that the system did not properly separate the words.

⁵<https://prodi.gy/>

⁶https://serviceportal.unibe.ch/sp?id=kb_article&view&sysparm_article=KB0010032

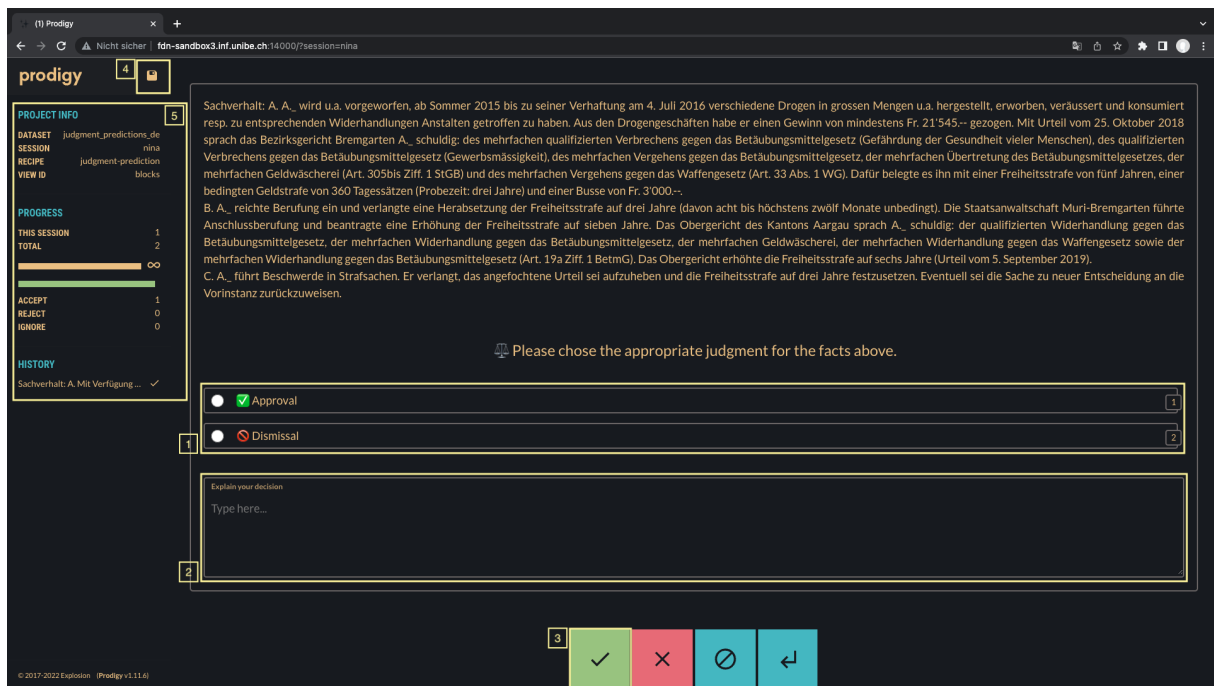


Figure 1: Screenshot of the case overview on prodigy

4.3 Reject or Ignore a Case

To reject a case state your reasoning in the comment section and press the red cross to reject it. To ignore it, press the blue button with the stop signal after commenting. Do not forget to save your progress. Figure 2 shows the interface of the comment section and the ignore and reject buttons.

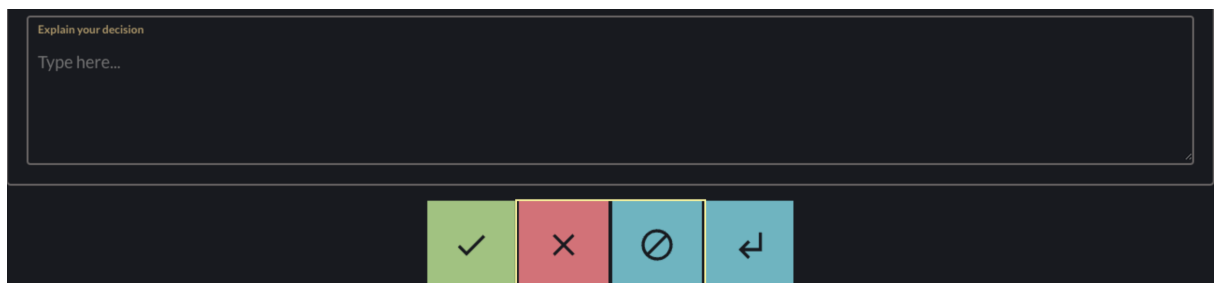


Figure 2: Reject and Ignore buttons

5 Change Log

This change log documents the progress of these guidelines. When adapting these guidelines please also add a new entry to the change log using the following structure

Template

Date – Title of changes

- Which parts were changed in this iteration?
- Why was this part changed?

References

- Leitner, E., Rehm, G., & Moreno-Schneider, J. (2019, 9). Fine-grained Named Entity Recognition in Legal Documents. In M. Acosta, P. Cudré-Mauroux, M. Maleshkova, T. Pellegrini, H. Sack, & Y. Sure-Vetter (Eds.), *Semantic systems. the power of ai and knowledge graphs. proceedings of the 15th international conference (semantics 2019)* (pp. 272–287). Karlsruhe, Germany: Springer. (10/11 September 2019)
- Niklaus, J., Chalkidis, I., & Stürmer, M. (2021). *Swiss-judgment-prediction: A multilingual legal judgment prediction benchmark*. arXiv. Retrieved from <https://arxiv.org/abs/2110.00806>
DOI: 10.48550/ARXIV.2110.00806
- Pustejovsky, J., & Stubbs, A. (2012). *Natural language annotation for machine learning* (Nos. Bd. 9,S. 878). O'Reilly Media, Incorporated.
- Reiter, N. (2020). Anleitung zur erstellung von annotationsrichtlinien. In N. Reiter, A. Pichler, & J. Kuhn (Eds.), *Reflektierte algorithmische textanalyse: Interdisziplinäre(s) arbeiten in der creta-werkstatt* (pp. 193–202). De Gruyter. Retrieved from <https://doi.org/10.1515/9783110693973-009>
DOI: doi:10.1515/9783110693973-009
- Wiegrefe, S., & Marasovic, A. (2021). Teach me to explain: A review of datasets for explainable NLP. *CoRR*, abs/2102.12060. Retrieved from <https://arxiv.org/abs/2102.12060>

List of Figures

1	Screenshot of the case overview on prodigy	4
2	Reject and Ignore buttons	4