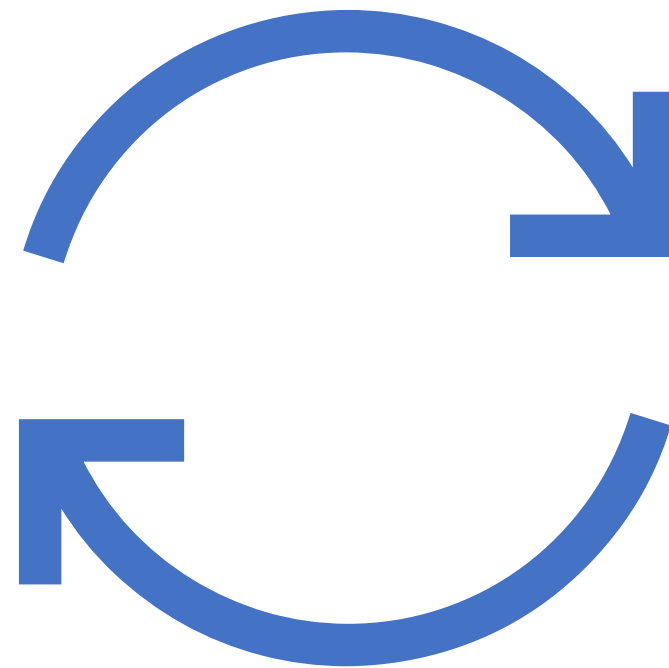


A Guide to understanding NER data

Annotations performed on SEBI
regulations and sub-regulations



Legal Doc: An entity referring to a standard, publicly available legal document including acts, regulations, legislatures, codes, clauses, sections, sub-sections, and Gazettes.

Object – General: An entity which is acted upon by authority, subject-Individual and subject-organization.

Subject-Individual: Subjects who are specific individuals or a class of individuals who engage in the securities market in any manner and upon whom, the acts, rules and regulations apply.

Subject-Organization: Companies or organizations, such as businesses or government departments, who are bound to act as per the rules, regulations, acts as well as directions of the authorities.

Authority: Entities which describe governing, regulatory, reporting, independent, decision making, authoritarian and rulemaking bodies of the rules.

Our Named entity tags

Value: An entity representing a literal numerical value.

Objective: An action which is the main theme or target of a regulation.

Transaction: An action that is not the main theme or target of a regulation but is ancillary to its fulfillment.

Legal Term: An entity representing official legal or financial terms relating to the legal or financial domain.

Date/Time: An entity representing mentions date and time in the text including numerical year, month, week, day, time, hour, and seconds as well as time/date in a relative to some event/occurrence.

Our Named entity tags

Annotated Data

A total of 189 sub-regulations were annotated by legal experts.

These contained about **1258** entity instances in them, out of which **134** (10.65%) overlapping entities were found.

A json file is present which has all these regulations annotated.

Frequency Counts for each Named entity

NER Label	Frequency
Objective	169
Object-General	124
Subject-Organization	90
Legal Term	82
Transaction	89
Subject-Individual	70
Legal Doc	67
Date/Time	35
Value	29
Authority	19

Data Description

JSON property name	Description
id	Identifier for the labeling task from the dataset.
data	Data copied from the input data task format
project	Identifier for a specific project in Label Studio.
annotations	Array containing the labeling results for the task.
annotations.id	Identifier for the completed task.
annotations.lead_time	Time in seconds to label the task.
annotations.result	Array containing the results of the labeling or annotation task.
annotations.completed_by	User ID of the user that created the annotation. Matches the list order of users on the People page on the Label Studio UI.
result.id	Identifier for the specific annotation result for this task.
result.from_name	Name of the tag used to label the region
result.to_name	Name of the object tag that provided the region to be labeled.
result.type	Type of tag used to annotate the task.
result.value	Tag-specific value that includes details of the result of labeling the task. The value structure depends on the tag for the label.
predictions	Array of machine learning predictions. Follows the same format as the completions array, with one additional parameter.
predictions.score	The overall score of the result, based on the probabilistic output, confidence level, or other.

Each regulation has several annotation completions. A completion has start, end positions, label name and text of the named entity labelled.

