Primary Approaches to Text-based Classification

There exist two prevailing types of solutions to text-based classification problems. One is a rules-based text classifica
fication.

In the rules-based text classification method, one employs the use of manually constructed logical rules to classify te
, such an
implementation would resemble an extensively built and potentially nested if-else statements [1]. Because of the ma
it is expected to
come with some maintenance overhead as business rules evolve over time.

For machine-learning-based text classification, one trains and tunes a supervised machine learning model to process
y bodies of text.

This path requires significant but non-repetitive upfront time and effort to build the training data set as well as fine-t
to classify bodies of text
[1]. However, it is expected to better handle the nuances of the language it has been trained on as compared to the

Typical Natural Language Processing (NLP) requires a preprocessing stage where the source raw bodies of text are c
ner that a machine learning algorithm can ingest and process [2, 3]. This is a necessary dependency for the machine
d approach.

Purpose of Rules-based Approach

The purpose of the Rules-based approach in the context of using a rule engine is to abstract away the implementatio
omplex if-else
statements and enables us to reconfigure the rules freely for sake of experimenting which is crucial for this initial sta
brary where we can
accomplish just this very goal in [5].

In the context of the use case of performing property and loan segmentation on our own data set, it is the first step
machine learning model
to take over the classification of properties and their associated loans.