

FH Kufstein Tirol

Web Business & Technology

Abstract of the thesis: **Advanced Text Document Classification Techniques:
Enhancing Accuracy and Efficiency in Real-World Applications**

Author: Jesse Lang

First reviewer: Prof. (FH) Dr. Michael Kohlegger

This thesis delves into the realm of advanced text document classification techniques, merging machine learning with natural language processing (NLP) to refine the accuracy and efficiency of categorizing diverse textual documents. The overarching goal is to discern the efficacy of such techniques in real-world applications and to unravel the factors influencing their performance.

Methodologically, the study embarks on a comprehensive journey. Beginning with data collection, it progresses through a series of text preprocessing steps, including tokenization, lowercase conversion, and stopwords removal, aimed at refining the data. Feature extraction techniques such as Bag-of-Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF) are then employed. The data undergoes a rigorous train-test-split process to facilitate model evaluation. Selection and training of classification models, spanning traditional machine learning to deep learning approaches, follow suit.

The culmination of this research yields compelling insights. Findings underscore the potency of advanced text document classification techniques in heightening accuracy and efficiency. Through meticulous experimentation and analysis, critical revelations emerge regarding model performance, data preprocessing intricacies, and runtime considerations. Noteworthy trends and discoveries pertaining to model performance and data characteristics surface, illuminating the landscape of text document classification.