

# Detect Job Dissatisfaction Among Nurses and

Degree programme : BSc in Computer Science  
Specialisation : Data Engineering  
Thesis advisor : Prof. Dr. Souhir Ben Souissi  
Expert : Dr. Andreas Spichiger

This thesis explores the application of machine learning and deep learning techniques to detect job dissatisfaction among nurses and healthcare professionals on social media. The focus is on developing, training, and evaluating NLP models to optimize job satisfaction classification performance.

## Introduction

Job dissatisfaction and high turnover rates remain persistent challenges in the healthcare sector. According to the Swiss Health Observatory, nearly one in four healthcare workers leaves their workplace each year in Switzerland this often unexpectedly. These sudden departures strain team dynamics and increase workload for the remaining staff. This bachelor thesis investigates the feasibility of using Natural Language Processing (NLP) to automatically detect job satisfaction in nursing-related Reddit posts.

## Goal

To explore whether job satisfaction can be detected from text based social media entries, in this work I collected and created a dataset with Reddit posts from nurse-related subreddits. Those posts were manually labeled with binary classification regarding job satisfaction.

In the next step, the study explored the use of pretrained models with customized Neural Network (NN) to improve performance. Additionally, Retriever-Augmented Generation (RAG) was investigated to enhance the model's ability to investigate specific reasons for job dissatisfaction.

## Methodology

The dataset originally contained 489 Reddit posts, which were later expanded to 555 posts. These posts were categorized into „satisfied“ and „dissatisfied“ based on job satisfaction. Augmentation techniques were used to extend the dataset further.

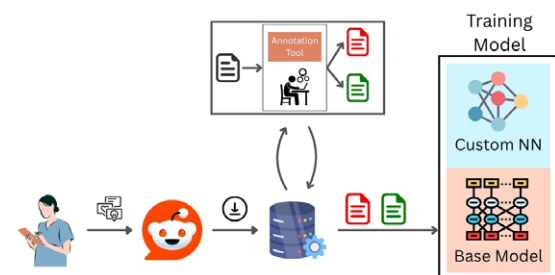
The dataset was then preprocessed and different classical embedding-based models and classifiers were implemented. This approach included models like Word2Vec, TF-IDF and Doc2Vec paired with classifiers such as SVM, Logistic Regression or Random Forest. Transformer-based models like BERT, RoBERTa and DeBERTa were implemented followed by a custom neural network to further optimize the performance.



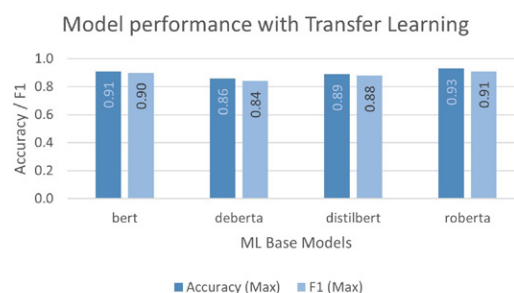
Marco Luca Ging

## Results

The results indicate that transfer learning and the state-of-the-art deep learning models are effective to detect job satisfaction. Among the models tested, RoBERTa achieved the best performance (Accuracy: 0.93, F1-score: 0.91), closely followed by BERT. DeBERTa and DistilBERT did not surpass the 0.9 accuracy threshold.



Overview trainig process



Results transfer learning: Acc/F1 according to base model