

Posudek školitele na diplomovou práci:

### **Active Learning for Text Classification**

Autor práce: **Bc. Marko Sahan**

Předložená práce se zabývá aplikací teorie aktivního učení na problém klasifikace textových dokumentů. Výchozím stavem je množina dokumentů, z nichž pouze malá část má známý správný výsledek klasifikace. Je však možné se na klasifikaci libovolného dokumentu dotázat anotátora. Cílem aktivního učení je minimalizovat počet dotazů potřebných k naučení co nejlepšího klasifikátoru. Tato úloha je netriviální kombinací teorii rozhodování, reprezentace neurčitosti a reprezentace textových dokumentů.

Práce je rozdělena do šesti hlavních kapitol. V první kapitole je představena teorie rozhodování za neurčitosti, ukázáno jak tato teorie vysvětluje základní algoritmy strojového učení a v neposlední řadě její aplikace na problém aktivního učení. V této kapitole je také představen algoritmus DEnFi, který je autorovou modifikací předchozího algoritmu. Nejnutnější pojmy z teorie zpracování textových dokumentů jsou představeny ve druhé kapitole. Ve třetí kapitole jsou představeny datové sady na kterých budou metody testovány a ve čtvrté kapitole implementace testovacího softwaru. Hlavní výsledky testů jsou prezentovány v páté a šesté kapitole. V páté kapitole jsou srovnány metody bez aktivního výběru anotovaných dokumentů, v šesté s aktivním výběrem. Hlavním výsledkem práce je prokázání účinnosti algoritmu DEnFi, který v testech překonal i stávající řešení založené na metodě Dropout.

Zadání diplomové práce bylo beze zbytku splněno. Ačkoliv algoritmus DEnFi vychází z předchozí práce, autor provedl a otestoval několik vlastních modifikací. Výsledná práce je napsána pečlivě a čitelně s malým množstvím překlepů a nejasností. Autor se podrobně seznámil jak s teoretickými metodami tak s reálnými daty. Autor pracoval intenzivně, metodicky a v oblasti reprezentace textu velmi samostatně. Výsledky diplomové práce byly zaslány k posouzení na konferenci COLING 2020.

Práci doporučuji k obhajobě, hodnotím ji známkou **A** – výborně.

Doc. Ing. Václav Šmídl, Ph.D.