

Tato práce se zabývá výzkumem v oblasti klasifikace dokumentů. Popisuje již existující techniky s důrazem na Naivní Bayesův klasifikátor. Zmíněny jsou i některé z metod pro výběr příznaků.

Teoretické pozadí je základem pro implementaci klasifikační knihovny založené na metodě Naivního Bayesovského klasifikátoru. Knihovna poskytuje kromě samotného klasifikátoru i paletu nástrojů pro předzpracování textu. Tyto nástroje umožňují práci s rozličným typem dokumentů, ale především značně snižují nadbytečné dimenze vstupních dat.

Knihovna je testována na dvou různých referenčních datových sadách na kterých jsou diskutovány rozdíly chování jednotlivých metod pro výběr příznaků. Funkčnost celé knihovny je prakticky ověřena jejím začleněním do open-source emailového klienta Mailpuccino.