

Posudek školitele na diplomovou práci:

Semisupervised learning of heterogenous structured data

Autor práce: **Bc. Michaela Mašková**

Předložená práce se zabývá problémem strojového učení nad heterogenními strukturovanými daty, tj. datovými záznamy různých typů uspořádanými ve stromové struktuře. Hlavním cílem je rozšířit existující metody vyvinuté pro učení klasifikátorů z dat s kompletní znalostí tříd na metody, které se dokáží využít dodatečná data bez známé třídy (tzv semisupervised učení). Tato oblast je značně nestandardní a práce byla i prací vědeckou.

Práce je rozdělena do pěti hlavních kapitol. V první kapitole jsou představeny obecné metody učení. V druhé kapitole jsou představeny specifika strukturovaných dat a standardní nástroj na jejich zpracování. Jádrem práce je aplikace metod učení z nekompletních dat na hierarchická strukturovaná data představená ve třetí kapitole. Ve čtvrté kapitole je popsán způsob testování na datech ze standardních databází a datech z praxe (poskytnutých firmou Avast). Výsledky jsou shrnuty v páté kapitole.

Autorka na práci pracovala systematicky a velmi samostatně. Většinu úkolů a zadání plnila rychle a nad rámec požadavku. Vysoce hodnotím i kvalitu kódu použitého k vygenerování výsledků, který je psán systematicky a je snadné jej upravit a rozšířit o další funkce. Experimentální vyhodnocení všech variant modelu bylo náročné a množství výsledků je pro diplomovou práci nadstandardní.

Samotná práce je napsána velmi dobrou angličtinou s minimem chyb a nejasností. Hlavním výsledkem práce je srovnání základních přístupů učení pro nový typ dat. Výsledky naznačují, že přínos dat bez známé třídy pro kvalitu učení je velmi omezený. Tento výsledek považuji za významný jako ukotvení základních metod, které je možné dále zlepšovat.

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

Doc. Ing. Václav Šmídl, Ph.D.