

Posudek vedoucího závěrečné práce

Název práce: **Neúplná Data a Rozhodovací Lesy v Úloze Klasifikace Šifrovaného síťového provozu**

Autor: **Lukáš Sahula**

Typ práce: Bakalářská práce

Katedra počítačů, Fakulta Elektrotechnická, ČVUT v Praze

Vedoucí: **Ing. Jan Brabec**

Pracoviště vedoucího: Katedra počítačů, Fakulta Elektrotechnická, ČVUT v Praze

Student se v rámci své práce zabýval technikami práce s neúplnými daty, které lze aplikovat na úlohu klasifikace malwaru ze šifrovaného síťového provozu. Práce byla zhotovena ve spolupráci se společností Cisco.

Zadání bych hodnotil jako náročnější z několika důvodů: Student pracoval s reálnými daty, která jsou v několika ohledech problematická. Jedná se o velká data, která nešla zpracovat na studentově osobním počítači, ale musel se s nimi naučit pracovat vzdáleně. I tak, student několikrát narazil na limity standardních nástrojů jako např. knihovny Pandas a bylo třeba řešit implementační problémy, se kterými by se student na menších úlohách nesetkal. Aby bylo možné evaluovat některé algoritmy pro práci s neúplnými daty, bylo nutné aby student implementoval vlastní verzi nahodných lesů. Student také musel implementovat komplikovanější evaluační algoritmy, aby byly výsledky práce relevantní. V rámci zadání softwarového projektu, na který navazuje tato práce, implementoval knihovnu v Pythonu zahrnující nástroje pro evaluaci a trénování klasifikátorů, která bude znovupoužitelná jak pro případnou navazující práci, tak pro další studenty.

V rámci práce byly evaluovány pouze dříve publikované algoritmy a na vlastní metody a vylepšení, která by mohla pomoci v případě konkrétní úlohy klasifikace šifrovaného provozu bohužel nezbyl čas. Nicméně, práce obsahuje poměrně užitečnou analýzu těchto konkrétních dat a ukazuje, které případné směry zpracování neúplných dat nedávají pro naše data smysl.

Zadání práce považuji za splněné ve všech bodech.

Student na bakalářské práci pracoval od srpna 2017. Pracoval pravidelně a docházel na konzultační schůzky skoro každý týden. Student je šikovný programátor, ale tato práce byla jeho první zkušenost se strojovým učením. Z toho důvodu, a pravděpodobně také vzhledem k náročnějšímu zadání, práce občas postupovala pomaleji než bych očekával. Také bych si dokázal představit výrazně více samostatné práce, iniciativy a nápadů ze studentovy strany.

Práce je psaná dobrou angličtinou a práce se zdroji je korektní. Délka textu činí skoro 40 stran, což by ji mělo řadit mezi delší bakalářské práce.

Po zahrnutí všeho výše zmíněného, s přihlédnutím k tomu, že se jedná o bakalářskou práci a tomu, že student pracoval poctivě a pravidelně, doporučuji předloženou práci k obhajobě a hodnotím ji klasifikačním stupněm: **B (velmi dobře)**

Datum: 3.6.2018

Podpis:

