Posudek diplomové práce pana Bc. Martina Bárty nazvané „Datamining výsledků experimentů“
Vedoucí práce doc. Dr. Ing. Tomáš Brandejský

Diplomant na základě své znalosti programování, operačních systémů Linux a Windows, databází a algoritmizace implementoval v jazyce Scala v prostředí Appache Spark určeném pro práci s Big data prostředí a algoritmy pro vyhodnocení výsledků velmi rozsáhlého souboru experimentů z oblasti genetického programování.

Práce z formálního hlediska odpovídá doporučené šabloně a čítá 83 stran včetně všech požadovaných seznamů. K práci přiložený ZIP soubor obsahuje vedle vlastního textu práce i zdrojové kódy, databázové soubory, vzorek vyhodnocovaných dat a Zeppelin sešity.
Hned v úvodu se mi líbila formulace o cestě (v buddhistickém smyslu), která asi nejlépe vystihuje celý problém dataminingu. Pracuje se s velkým objemem dat, který přesahuje možnosti počítače a běžných databázových a analytických nástrojů. K tomu jsou využívány komplexní nástroje, jejichž zvládnutí není otázkou minut, nebo dnů. Přesto je třeba mnoho věcí doprogramovat buď v běžných jazycích (Java, Python), ale s mnoha omezeními, nebo ve specializovaných jazycích jako Scala (byl použit v této práci), ale ty je třeba zvládnout, což opět není triviální problém. Scala kombinuje objektové a procedurální programování. K tomu je třeba zvolit několik vhodných knihoven, ale na výběr jsou jich doslova tisíce. Na víc ne všechny kombinace knihoven jsou použitelné. To vede k paradoxu, že programy pro analýzu Big dat bývají zdánlivě jednoduché, ale za jejich vytvořením stojí velký objem práce s pochopením nejen použitých, ale i nakonec nepoužitých nástrojů a knihoven a jejich konfigurací. Také vlastní vývoj probíhá poměrně pomalu vzhledem k výpočetním nárokům úloh, které ani pro potřeby vývoje nelze neomezeně zjednodušovat.
Práce je strukturována přehledně a logicky členěna do jednotlivých kapitol. Začíná problematikou dolování dat, které bylo v tomto případě nahrazeno neméně pracným výpočtem numerických experimentů, které proběhly v uplynulých letech nejen na našem clusteru v Upce, ale také na systémech Metacentra s superpočítači Anselm v Ostravě. Poté student představil použité technologie (a pro stručnost zamlčel řadu nakonec nepoužitých). V další kapitole popsal systém Sparc pro práci s Big daty a srovnal jej i s jeho předchůdcem Hadoop. V obsáhlé čtvrté kapitole popsal přípravu prostředí – tedy nalezení konzistentní množiny nástrojů pro řešení zadaného problému a jeich provázání. Právě tyto zkušenosti jsou stejně cenné, jako řešení zadaného problému – implementace dataminingového nástroje. V páté kapitole diplomant rozebírá vlastní analýzu vzorových dat včetně nalezených výsledků.
Student pracoval samostatně, projevoval značnou iniciativu a odvedl značný objem práce. Sám se seznámil s velkým množstvím nástrojů a knihoven.
Zadané cíle práce byly beze zbytku splněny. Po typografické stránce je práce zdařilá. Diplomant se pouze nevyvaroval drobných překlepů, např. „Exspirované“ str. 67, popisek obr. 26.
Kontrola plagiátorsví zřejmě díky nepříliš frekventovanému tématu, použitým knihovnám a neobvyklému programovacímu jazyku uvedla neuvěřitelnou míru shody 0%, proto můžeme práci považovat za zcela původní.
Práci považuji za zdařilou, velmi obsáhlou a zabývající se na diplomovou práci velmi rozsáhlou problémovou oblastí, kterou se z pohledu nároků na diplomovou práci podařilo diplomantovi zvládnout.
Proto navrhuji diplomovou práci Bc. Martina Bárty doporučuji k obhajobě a hodnotím ji vzhledem k její neobvyklosti, rozsahu a náročnosti známkou A, tedy Výborně.

V Pardubicích 4.9.2020

doc. Dr. Ing. Tomáš Brandejský