

**Jméno bakaláře:** Jan Kandyba

**Garantující katedra:** KKY

**Název bakalářské práce:** Detekce tématu dokumentu

	Předmět hodnocení	Nadprůměrné	Průměrné	Podprůměrné
1	Jazyková a grafická úprava	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
2	Samostatnost zpracování tématu BP	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3	Vhodnost použitých metod	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	Způsob zpracování a vyhodnocení	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
5	Správnost získaných výsledků	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6	Vlastní přínos	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Doplnění hodnocení, připomínky, dotazy:

Student bakalářského studijního programu touto prací plně reprodukoval výsledky vědeckého článku vzniklého na katedře kybernetiky v předchozím roce, zabývajícím se detekcí tématu v telefonních hovorech jazykové poradny Ústavu pro jazyk český Akademie Věd ČR. V rámci této práce se seznámil s problematikou zpracování přirozeného jazyka a state-of-the-art metodami pro popis a klasifikaci dokumentů. Rozsah znalostí potřebných k splnění zadaného úkolu převyšoval nabitě znalosti z bakalářského studia, student však prokázal schopnost převážně samostudiem tyto znalosti sám vstřebat a aplikovat.

Drobné výhrady bych měl k formulaci některých odstavců v textu samotné bakalářské práce, ale s přihlédnutím na komplexnost neznámého problému, se kterým se musel student seznámit, mi tento nedostatek přijde jako marginální.

Nad rámec práce student otestoval různá nastavení neuronových sítí a jejich využití pro klasifikaci. S ohledem na rozsah vykonané práce se přikláním k hodnocení - výborně.

Otázky:

Co je podtrénování a přetrénování klasifikátoru (v práci vysvětleno ve spojitosti s délkou trénování)?

Jak postupovat při klasifikaci problému se silně nevyváženými třídami (např. detekce vzácných onemocnění)? Jak bychom museli upravit naši úlohu, pokud bychom chtěli detekovat i ty třídy, které jsou v datech málo zastoupeny?

Proč padla volba na ztrátovou funkci založenou na kategoričké křížové entropii při trénování neuronové sítě?

Splnění bodů zadání	<input checked="" type="checkbox"/> úplně	<input type="checkbox"/> částečně	<input type="checkbox"/> nesplněno
Doporučení práce k obhajobě	<input checked="" type="checkbox"/> ano	<input type="checkbox"/> ne	
<b>Celkové hodnocení práce</b>	<input checked="" type="checkbox"/> <b>výborně</b>	<input type="checkbox"/> <b>velmi dobře</b>	<input type="checkbox"/> <b>dobře</b> <input type="checkbox"/> <b>nevyhověl</b>
Jméno, příjmení, titul vedoucího BP: Ing. Zbyněk Zajíc, Ph.D.			