

МОЛБА

Школска 2023/24. година

ПОДАЦИ О ПОДНОСИОЦУ МОЛБЕ

Презиме:	Ђорђевић	Име:	Петар
Е-пошта:	mi221088@alas.matf.bg.ac.rs		
Број индекса:	1088/2022	Статус:	Самофинансирање
Ниво студија:	Мастер академске студије		

Тип: пријава теме мастер рада

Број молбе: m1088/2022-2

Катедра: Катедра за рачунарство и информатику

Развој ефикасног и скалабилног тока података помоћу Apache Airflow-а и Spark-а

САДРЖАЈ МОЛБЕ:

Значај теме и области:

Област овог рада обухвата инжењерство података, са посебним фокусом на уређивање токова података. Инжењерство података и токови података имају значајан утицај на обликовање инфраструктуре савременог пословног окружења. Главни циљ токова података је да омогуће ефикасно прикупљање, складиштење, анализу, обраду и коришћење великих количина података, уз могућност прилагодљивости новим или промењеним захтевима и аутоматизацији процеса како би се осигурала конзистентност и ефикасност у управљању подацима.

Специфични циљ рада:

Циљ овог рада је да се истраже и упознају све фазе процеса током управљања подацима, од почетног прикупљања па све до крајњег тренирања модела. Акценат је на разматрању могућности решења отвореног кода, а посебно:

1. Apache Airflow за оркестрацију, аутоматизацију и визуелизацију тока података
2. Apache Spark са Apache Iceberg форматом складишта за ефикасно и скалабилно процесирање велике количине података
3. MLFlow као регистар модела за праћење, управљање и репродукцију експеримената са моделима.

Као део овог рада биће развијен пројекат који кроз микросервисну архитектуру демонстрира примену описаних алата и савремених пракси у области инжењерства података, који ће моћи да послужи као пример употребе и шаблон ширем кругу корисника. Конкретни ток података који ће бити имплементиран је анализа филмова. Овај ток података прикупља информације о филмовима из базе података као што је IMDb, укључујући податке о режисерима, главним глумцима, оценама, буџету и заради. Затим, трансформацијом и обogaћивањем ових података, врше се процесирање и анализа који могу да пруже увид у популарност и зараду филмова и факторе који утичу на њихову привлачност публици. На основу ових података тренираће се модели који покушавају да предвиде наведене факторе.

Комисија:

- 1. Ментор: Саша Малков
- 2. Председник: Иван Чукић
- 3. Члан: Мирјана Маљковић Ружичић

ПРИЛОЗИ МОЛБИ:

- 1. образац молбе за одобрење теме мастер рада

Београд, 21. мај 2024.

(потпис)

(Попуњава надлежна особа)

☒ **ОДОБРАВА СЕ**

☐ **НЕ ОДОБРАВА СЕ**

21. мај 2024., Мирослав Марић

(датум, име, презиме и потпис)

ОБРАЗЛОЖЕЊЕ: