

ИЗВЕШТАЈ О ОЦЕНИ МАСТЕР РАДА

I ПОДАЦИ О КОМИСИЈИ
<p>1. Датум и орган који је именовao Комисију 28. августа 2024. Веће Департмана за математику и информатику Природно-математичког факултета Универзитета у Новом Саду</p> <p>2. Састав Комисије са назнаком имена и презимена сваког члана, звања, назива уже научне области за коју је изабран у звање, датума избора у звање и назив факултета, установе у којој је члан комисије запослен:</p> <ul style="list-style-type: none">• Др Душан Јаковетић, ванредни професор ПМФ-а у Новом Саду, ужа научна област: математичко моделирање, изабран у звање: 15.11.2020. године, председник• Др Оскар Марко, научни сарадник Института БиоСенс у Новом Саду, ужа научна област: електроника, телекомуникације и информационе технологије, изабран 22.01.2020. године, ментор• Др Марко Панић, научни сарадник Института БиоСенс у Новом Саду, ужа научна област: техничко – технолошке науке - информационе технологије, члан
II ПОДАЦИ О КАНДИДАТУ
<p>1. Име, име једног родитеља, презиме: Наташа, Зоран, Диклић</p> <p>2. Датум рођења, општина, република: 10.11.1994, Прњавор, Република Српска, БиХ</p> <p>3. Година уписа на дипломске академске студије, смер/усмерење: 2023/24, Мастер математичар, Примењена математика (модул: наука о подацима)</p>
III НАСЛОВ МАСТЕР РАДА
„Предвиђање симптома изазваних алергијом на полен коришћењем алгоритама машинског учења“
IV ПРЕГЛЕД МАСТЕР РАДА
Мастер рад „Предвиђање симптома изазваних алергијом на полен коришћењем алгоритама машинског учења“ се састоји од 55 страна. Садржај је распоређен у 5 поглавља: увод, материјали и методи, резултати, дискусија и идеје за будући рад, и закључак. Рад почиње апстрактном, захвалницама, садржајем, затим је дат списак фигура. Рад садржи 34 фигуре. У раду је представљен опис, анализа и препроцесирање података, математичка теорија која је фундамент алгоритама машинског учења, опис коришћених алгоритама и експериментални резултати добијени након тренирања и валидирања алгоритама. На крају је дат преглед литературе која је коришћена од укупно 30 референци.
V ВРЕДНОВАЊЕ ПОЈЕДИНИХ ДЕЛОВА МАСТЕР РАДА
<u>Прво поглавље - Увод</u> Теза почиње уводом у основне концепте који су основа овог истраживања и представља мотивацију за настанак овог рада. Говори се о алергијама на полен и

њиховом предвиђању уз помоћ машинског учења. Објашњава се како полен утиче на људе и колико је важно прецизно предвидети симптоме алергије како би се смањили негативни ефекти на свакодневни живот. Затим се уводи појам машинског учења, описују се два главна типа машинског учења (надгледано и ненадгледано), наводе се алгоритми: Random Forest, Gradient Boosting, XGBoost, SVC и K-Nearest Neighbors. Указује се на значај примене машинског учења у области аеробиологије, са посебним освртом на научне радове који су се бавили сличним истраживањима. Ови радови укључују предикције симптома повезаних са алергијом на полен, као и предикције концентрације полена у ваздуху, што су кључни аспекти за боље разумевање и управљање сезонским алергијама.

Друго поглавље – Материјали и методи

У овој студији коришћени су подаци из два извора: Дневник пацијената са поленском алергијом (Patients' Hayfever Diary) и дневна мерења полена Лабораторије за Палинологију, ПМФ Нови Сад. Податке је обезбедио др. Б. Шикопарија из Института BioSense, који је прикупио мерења полена коришћена у овом раду. Подаци о симптомима су добијени у сарадњи са ОРЛ одјељењем Медицинског универзитета у Бечу. Скуп података "PHD-2009-2017.csv" садржи информације о корисницима дневника и њиховим симптомима и медицинским третманима, док "ParticleCountDailyVojvodina2009-2017.xlsx" садржи податке о дневном броју поленских честица за период од 2009. до 2017. године. Анализирани су подаци из оба скупа података. Истражени су хистограми уноса података по корисницима, корелације између симптома и концентрација полена, као и варијације симптома током године. Презентоване су корелације између симптома, месечне варијације симптома и употребе медицинских третмана. Показано је да највиша концентрација полена пада у пролеће и јесен. Одрађено је препроцесирање података које укључује мапирање два скупа података по датумима, избор атрибута, попуњавање недостајућих вредности и груписање података. Анализиране су две методе за предвиђање симптома алергије: предвиђање интензитета укупних симптома (од 0 до 21) и предвиђање типа симптома (очи, нос, плућа или њихове комбинације). За први приступ коришћена су четири регресиона модела: KNN, Random Forest, Gradient Boosting и XGB регресор. За други приступ испитано је пет класификационих модела: KNN, SVC, Random Forest, Gradient Boosting и XGB класификатор. Модели су обучени користећи дневне и кумулативне концентрације полена за периоде од 2, 3, 4, 5, 10 и 15 дана. Имплементација је урађена у програмском језику Python у Jupyter Notebook окружењу, уз коришћење библиотека као што су Scikit-learn, Pandas, Matplotlib и Datetime. Детаљно су објашњени коришћени алгоритми машинског учења. Такође, разматран је Grid Search као техника за оптимизацију хиперпараметара модела, где се користе различите комбинације параметара за побољшање перформанси. За евалуацију регресионих модела коришћене су метрике као што су средња апсолутна грешка (MAE) и корен средње квадратичне грешке (RMSE), док су за класификационе моделе коришћени тачност, прецизност, осетљивост и F1 скор.

Треће поглавље – Експериментални резултати

У овом поглављу су приказани експериментални резултати добијени применом различитих алгоритама машинског учења за класификацију и регресију, као и њихово поређење. Резултати су прво приказани након тренирања класификационих модела на подацима са инпутима који се понављају, затим на груписаним подацима, а на крају на подацима корисника који је највише користио полен дневник. Исти приказ резултата примењен је и за регресионе моделе за предвиђање интензитета укупних симптома.

Четврто поглавље – Дискусија и будући рад

Поглавље о дискусији тумачи добијене резултате у ширем контексту постојећих истраживања. Такође, предложене су идеје за будући рад. Анализом је утврђено да модели машинског учења постижу боље резултате на груписаним и персонализованим подацима, док понављајући уноси смањују тачност предвиђања. Предложени су правци за унапређење модела, укључујући побољшање квалитета података, унапређену оптимизацију хиперпараметара, проширење скупа карактеристика и примену временских модела. Будући рад би могао укључити развој персонализованих модела и њихову примену у реалном времену кроз корисничке апликације.

Пето поглавље – Закључци

У петом поглављу су описани закључци овог истраживања, где је резимирано истраживање и финални резултати.

VI ЗАКЉУЧЦИ ОДНОСНО РЕЗУЛТАТИ ИСТРАЖИВАЊА

Неколико закључака се може извући из анализе резултата:

Утицај понављајућих уноса на перформансе модела

Модели обучени на подацима са понављајућим уносима (идентични датуми и концентрације полена, али различити симптоми корисника) показали су лошије резултате. На пример, класификатор Random Forest је постигао тачност од само 0.49, што указује на високу стопу погрешних предвиђања. Регресиони модели такође су имали високе грешке.

Ефективност јединствених уноса

Модели обучени на подацима са јединственим уносима значајно су побољшали перформансе, где је XGB класификатор постигао тачност од 0.894. Регресиони модели су такође показали боље резултате са нижим вредностима средње апсолутне грешке (MAE) и корен средње квадратичне грешке (RMSE), нарочито када су се подаци акумулирали на дуге периоде. Регресори K-Nearest Neighbors, Random Forest и Extreme Gradient Boosting показали су побољшане перформансе са растућим кумулативним концентрацијама полена ($n = 1$ до 15 дана), што сугерише да агрегирање података током одређеног периода помаже у ефикаснијем хватању основних образаца.

Перформансе модела на индивидуалним корисницима

Модели обучени на подацима корисника са највише уноса (ID 16796) имали су боље резултате у односу на моделе обучене на подацима свих корисника. Оба модела Random Forest и Gradient Boosting су постигла високу тачност (0.83), док су у задацима регресије показали најниже вредности грешака (MAE од 2.11 и RMSE од 1.05).

Поређење перформанси различитих алгоритама

Енсамбл методе, као што су Random Forest, Gradient Boosting и XGB, показале су се супериорним у односу на једноставније моделе попут K-Nearest Neighbors, како у задацима класификације, тако и регресије.

VII КОНАЧНА ОЦЕНА МАСТЕР РАДА

VIII ПРЕДЛОГ

Нови Сад,
.2024.

ПОТПИСИ ЧЛАНОВА КОМИСИЈЕ

Проф. Др Душан Јаковетић, ванредни
професор ПМФ, председник

Проф. Др Оскар Марко, научни сарадник
Института БиоСенс, ментор

Проф. Др Марко Панић, научни сарадник
Института БиоСенс, члан
