

Fakulta informatiky a informačných technológií

STU v Bratislave

IMS-RHPT

Dokumentácia k inžinierskemu dielu

Členovia tímu: Gáborčíková Zuzana
Chalás Filip
Janec Marek
Kavoň Rastislav
Miklovič Matúš
Nemček Peter
Valkovičová Katarína

Vedúci tímu: Ing. Richard Marko, Phd., Mgr. Martin Sabo, Phd.

Email: tp.tim2@gmail.com

Semester: zimný 2021/2022

1. Úvod	2
2. Globálne ciele pre ZS	3
3. Celkový pohľad na systém	4
3.1. Zjednodušená architektúra systému	4
3.2. Dátový model	5
4. Moduly systému	6
4.1. Analýza webovej aplikácie	6
4.1.1. Technické požiadavky	6
4.1.2. Požiadavky na webovú aplikáciu	6
4.1.3. Prípady použitia	9
4.2. Návrh webovej aplikácie	19
4.2.1. Návrh homepage	19
4.2.2. Návrh administrátorského rozhrania	21
4.2.3. Návrh používateľského rozhrania	25
4.3. Implementácia webovej aplikácie	32
4.4. Testovanie webovej aplikácie	32
4.5. Analýza dátového modelu	32
4.5.1. Úvod do problematiky	32
4.5.2. Súvisiace práce	33
4.5.3. Základné informácie o dátach	34
4.6. Návrh riešenia dátového modelu	34
4.7. Implementácia dátového modelu	35
4.7.1. Spracovanie .txt súborov a vytvorenie .csv	35
4.7.2. Unifikácia dát	36
4.7.3. Prvotná vizualizácia dátového súboru	36
4.7.4. Odstránenie šumu v dátach	37
4.7.5. Agregácie meraní do jedného riadku	38
4.7.6. Porovnávanie vektorov	38
4.7.7. Výber črt	39
5. Referencie	40

1. Úvod

V našom projekte sa venujeme spracovaniu dát zo spektrometra a predikcii ideálneho času pre zber technického konope na základe získaných dát. V tomto dokumente opisujeme priebeh projektu. Definujeme globálne ciele zimný semester, celkový pohľad na webovú aplikáciu určenú pre pestovateľov technického konope. V krátkosti sú opísané funkcionality, ktoré bude aplikácia poskytovať.

2. Globálne ciele pre ZS

Na začiatku tímového projektu sme definovali globálne ciele pre zimný semester:

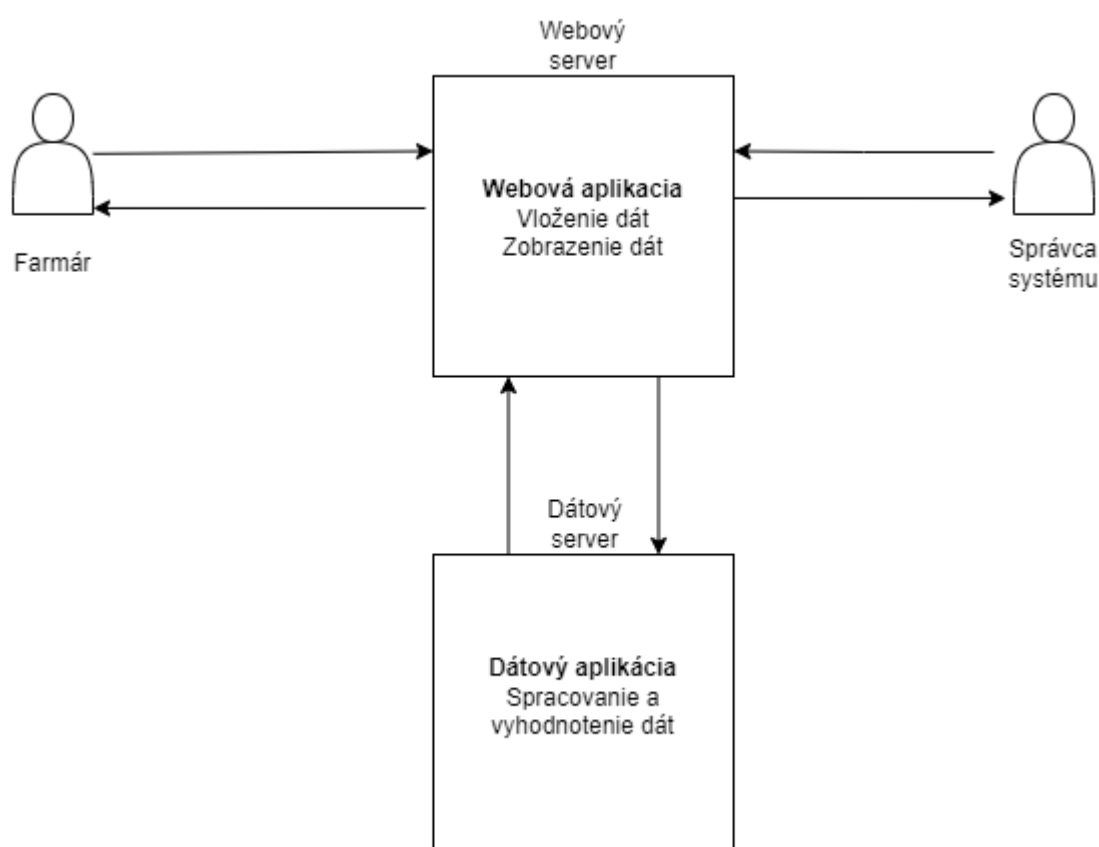
- Vedieť identifikovať hodnoty CBD zo spektra s čo najväčšou presnosťou.
- Mať prvú verziu funkčného modelu, ktorý bude vedieť predikovať počet dní chýbajúcich do zberu technickej konopy. Model bude pripravený na ďalšie vylepšovanie a optimalizovanie s cieľom dosiahnuť lepšiu presnosť.
- Mať dokončenú prvú verziu funkčnej webovej stránky, na ktorej sa bude dať zaregistrovať, prihlásiť a nahrávať dáta vo formáte txt alebo csv s tým, že budú používateľovi poskytnuté výsledky z modelu.
- Prepojiť model s webovou stránkou.

3. Celkový pohľad na systém

Výsledkom našej práce bude aplikácia, ktorá bude mať niekoľko hlavných funkcionalít. Sú nimi hlavne:

- Nahranie dát na server
- Zobrazenie výsledkov
- Spracovanie a vyhodnotenie dát

3.1. Zjednodušená architektúra systému



Aplikácia sa bude skladať z dvoch hlavných modulov. Sú nimi webová a dátová časť.

Webová časť bude slúžiť primárne ako prostredie pre používateľov - farmárov. Tí budú môcť do aplikácie nahrávať .txt súbory vygenerované spektrometrom.

Budú mať povytvárané vlastné profily, o ktoré bude možné požiadať na základe registračného formulára. Do aplikácie bude mať prístup aj správca, ktorého úlohou bude správa profilov, pridávanie nových používateľov.

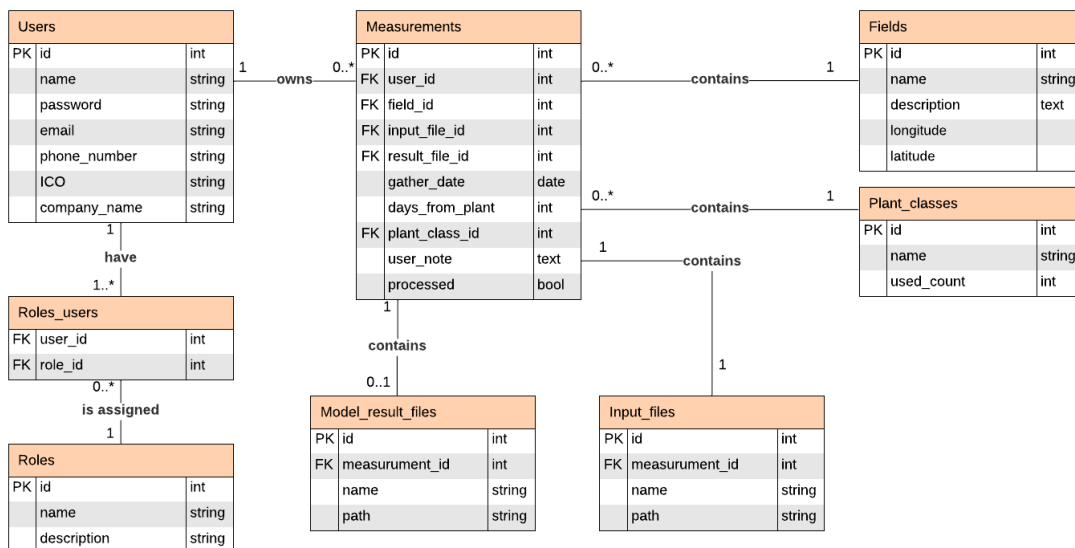
Vo svojom osobnom profile budú používatelia vidieť výsledky dát a rôzne iné štatistiky.

Dátová časť aplikácie bude slúžiť na výpočet a vyhodnotenie dát. Bude komunikovať s webovým serverom. Tieto dve časti aplikácie budú nezávislé. Komunikovať budú prostredníctvom vytvoreného API.

Úlohou dátovej časti bude na základe vytvoreného modelu spracovať súbor, ktorý dostane od webovej aplikácie. Po vhodnutí dát bude následne posilať výsledky naspäť.

3.2. Dátový model

Dátový model našej aplikácie obsahuje tabuľky so stĺpcami podľa nasledujúceho diagramu.



Tabuľka users slúži na ukladanie používateľov a zároveň aj žiadostí o registráciu. Pokiaľ ide o žiadosť, je to entita používateľa bez hesla, takže sa nemôže ešte prihlásiť dovedy, kým mu nie je žiadosť schválená a teda aj pridelené heslo. Až po pridelení hesla sa z entity stáva plnohodnotný používateľ. Deaktivácia používateľov bude zabezpečená pomocou predvoleného stĺpca - "deleted_at". Tabuľka rolí definuje role v systéme a pomocou prepojovacej tabuľky medzi rolami a používateľmi priraduje role jednotlivým používateľom. Role, ktoré v našom systéme vystupujú sú: admin pre administrátora a user pre bežného používateľa.

Tabuľka meraní v sebe obsahuje všetky potrebné väzby na ostatné entity tak, aby v sebe obsahovala všetko podľa definovaných požiadaviek. Konkrétne obsahuje väzby na tabuľku polí, druhov rastlín, vstupných súborov a takisto aj výstupných súborov. Entita merania môže existovať aj bez výstupného súboru. Musí však obsahovať pole, druh rastliny a vstupný súbor. Existencia entity merania bez výstupného súboru nastane po jej importovaní do systému, kedy ešte neexistuje výstupný súbor z modelu. Ten je k entite priradený až po spustení a výpočte výsledkov modelu. Polia a takisto aj druhy rastlín môžu byť priradené ku viacerým meraniam.

4. Moduly systému

4.1. Analýza webovej aplikácie

Pri analýze webovej aplikácie sme najskôr v spolupráci s product ownerom zadefinovali požiadavky, ktoré sme neskôr formálne spísali. Funkcionálne požiadavky, ktoré sú označené ako požiadavky na webovú aplikáciu sme následne rozpracovali do jednotlivých Use caseov - používateľských scenárov.

4.1.1. Technické požiadavky

Databáza

- V databáze budú uložené dáta o zaregistrovaných používateľoch, ako sú napr.: používateľské meno, heslo a pod.
 - meno a priezvisko
 - email (môže byť náhradou používateľského mena)
 - heslo (na začiatku by mohlo byť vygenerované a potom si ho bude môcť zmeniť)
 - firma (názov firmy)
- V ďalšej tabuľke budú uložené cesty k súborom, ktoré budú výstupmi merania (vstup do modelu), pričom každý súbor bude patriť jednému meraniu.
- Výsledky modelu budú uložené do súboru, ktorého cesta bude priradená k meraniu
- Merania používateľov sú uložené vo vlastnej tabuľke s údajmi (používateľ, mapovanie na pole - voliteľné, druh rastliny (druh len ako string pre používateľa), mapovanie na súbory patriace k meraniu (vstup aj výstup modelu), dátum zberu vzorky, časový údaj od zasadenia po odobratie vzorky v dňoch)
- Vlastná tabuľka pre polia, ktoré budú priradené k používateľom, a ich atribúty budú názov, prípadne opis a geolokácia
- Žiadosti o registráciu (meno, email, telefón (voliteľné), IČO (voliteľné), názov firmy, stav žiadosti)

Responzivita

- Používateľské rozhranie bude prispôsobené rôznym zariadeniam, teda rôznym veľkostiam obrazovky

4.1.2. Požiadavky na webovú aplikáciu

Homepage

- Na hlavnej stránke budú informácie ohľadom tohto projektu, na čo slúži a ako môže pomôcť poľnohospodárom pri pestovaní technickej konopy.
- V niekoľkých bodoch vysvetlíme, čo je náplňou projektu a ako funguje.
 - Meranie vzoriek iónovým spektrometrom
 - Čo je to iónový spektrometer a ako funguje
 - Model na dátach zo spektrometra - ako dokáže pomôcť

- Obsah hlavnej stránky by mal nalákať nezaregistrovaných používateľov poslať žiadosť o registráciu.

Prihlásenie používateľa

- Používateľ bude schopný prihlásiť sa na základe svojich prihlasovacích údajov (meno/email a heslo).

Registrácia

- Proces registrácie sa začne žiadosťou používateľa o registráciu. Ak ju používateľ s administrátorskými právami potvrdí, používateľovi príde na zadaný email heslo priradené k jeho účtu
 1. Používateľ pošle žiadosť cez formulár na webovej stránke a uvedie o sebe základné informácie
 2. Administrátorovi príde upozornenie do systému o zaevidovaní žiadosti
 3. V detaile žiadosti ju administrátor potvrdí alebo zamietne
 4. V prípade potvrdenia sa používateľovi vygeneruje heslo
 5. V oboch prípadoch sa odošle mail na uvedenú mailovú adresu s informáciou o potvrdení alebo zamietnutí žiadosti
 6. Ak je žiadosť potvrdená, súčasťou potvrdzovacieho emailu bude aj heslo pre používateľa (v emaili bude vyzvaný k zmene hesla po prihlásení)

Vkladanie dát konkrétnemu používateľovi

- Používateľ s administrátorskými právami bude môcť nahrať výsledky z merania (dáta) konkrétnemu používateľovi.
- Po prihlásení do systému bude môcť používateľ nahrať dáta z merania.
- V oboch prípadoch prebehne na pozadí kontrola formátu dát, či sú vhodné do natrénovaného modelu. Ak áno, vkladanie prebehne bez problémov. Ak tieto dáta nebudú vo vhodnom formáte, systém na to používateľa upozorní a nepovolí mu ich vložiť do systému.

Pridávanie polí

- Používateľ si bude môcť v systéme pridávať a spravovať polia, ktoré bude môcť neskôr z dropdownu vybrať pri pridávaní merania.

Druhy odrôd

- V administrátorskom rozhraní bude editor, kde bude možné pridávať a mazať jednotlivé druhy odrôd rastlín. Mazať druhy rastlín bude možné iba ak tento druh zatiaľ nie je použitý pri žiadnom meraní.

- Farmár si môže ručne doplniť odrodu, ktorá ešte nefiguruje v systéme pri pridávaní merania. Pridanie bude fungovať tak, že si nevyberie žiadny druh z dropdownu, ale zadá vlastný textový reťazec. Odroda sa automaticky doplní do zoznamu odrôd pre všetkých používateľ.

Spustenie modelu na dátach používateľa

- Po tom, čo používateľ úspešne vloží dáta do modelu, mu bude umožnené spustiť na týchto dátach predikciu regresného modelu.

Detail merania

- Každé meranie má špecifikované:
 - Odroda/druh konopy
 - Časový údaj od zasadenia po odobratie vzorky v dňoch
 - Miesto zberu/pole (každý používateľ môže mať definované vlastné polia s názvom, aby mohol merania agregovať podľa tohto atribútu)
 - Výsledky merania (napr. vizualizácia výsledkov zo spektrometra)
 - Výsledky z modelu
 - časový údaj, kedy treba zbierať
 - aktuálne množstvo CBD (toto asi nebude výsledok modelu)

Vizualizácia meraní

- Používateľ bude mať možnosť spojiť viac meraní do jednej vizualizácie, aby mohol vidieť zmenu v čase.
- Prípadná rozšírená možnosť (ak bude na to priestor a čas) - Súčasťou vizualizácii bude aj mapa Slovenska, na ktorej zobrazíme body (podľa GPS súradníc), z ktorých má používateľ namerané vzorky.

Práca s výsledkami

- Používateľ si bude môcť písať poznámky ku každému meraniu. Pri každom detaile merania môže používateľ písať poznámky.

Detail používateľa

- Administrátor si bude môcť prezrieť zoznam všetkých používateľov s možnosťou zobrazenia detailu používateľa. Pri detaile bude možnosť aj deaktivovať účet, pričom sa tento účet nevymaže a ani jeho údaje sa nevymažú. Deaktiváciou sa len znemožní prihlásenie.

4.1.3. Prípady použitia

Prípady použitia opisujú jednotlivé činnosti, ktoré budú môcť používatelia v rámci produktovej webovej aplikácie vykonávať. Vychádzajú z predchádzajúcej časti - Požiadavky na webovú aplikáciu, pričom je ich úlohou naplniť všetky požiadavky na aplikáciu.

V rámci prípadov použitia vystupujú nasledovné typy aktérov:

Používateľ – všeobecný používateľ

Farmár – všeobecný pestovateľ, ktorý sa môže dostať na stránku

Prihlásený farmár – farmár, ktorý má vytvorený účet na stránke a je autentifikovaný systémom

Administrátor – správca systému, ktorý má špeciálne práva, pre prihlásenie musí byť autentifikovaný

UC01 – Žiadosť o registráciu

Ako farmár chcem mať možnosť požiadať o registráciu a vytvorenie konta v systéme, aby som sa mohol neskôr prihlásiť a mať prístup k funkcionalitám, ktoré systém ponúka.

Predpokladom je, že farmár sa nachádza na home page webového rozhrania systému.

1. Farmár klikne na tlačidlo - "Žiadosť o registráciu"
2. Systém zobrazí formulár so žiadosťou o registráciu
3. Farmár vyplní potrebné údaje - (emailový kontakt, nejaké farmárske údaje?, prihlasovacie meno? možno heslo?...)
4. Farmár klikne na tlačidlo - "Odoslať žiadosť o registráciu"
5. Systém zaeviduje žiadosť o registráciu a pridá ju medzi žiadosti v databáze systému
6. Systém presmeruje používateľa na home page webového rozhrania

UC02 – Vytvorenie používateľského konta

Ako administrátor chcem vedieť vytvoriť používateľské konto používateľovi, ktorý o to požiadal prostredníctvom žiadosti o registráciu, aby som mu sprístupnil vlastné konto, v ktorom bude mať možnosti na využívanie systému.

Predpokladom je prihlásenie v administrátorskom rozhraní - na úvodnej stránke tohto rozhrania a existujúca podaná žiadosť o registráciu farmára.

1. Administrátor klikne na tlačidlo - "Žiadosti"
2. Systém zobrazí zoznam všetkých nevybavených žiadostí o registráciu
3. Administrátor klikne na jednu zo žiadostí v zozname - zvolí žiadosť, ktorú chce vybaviť
4. Systém zobrazí detail žiadosti spolu so všetkými jej informáciami
5. Administrátor skontroluje informácie v žiadosti a klikne na tlačidlo - "Vytvoriť používateľské konto"

6. Systém vytvorí konto a odošle prihlasovacie údaje na zadaný email v žiadosti
7. Systém označí žiadosť v databáze za vybavenú
8. Systém presmeruje používateľa do zoznamu všetkých nevybavených žiadostí o registráciu

Alternatívny scenár (od kroku 5) - nevyhovenie žiadosti

Predpoklad - nesprávnosť údajov v žiadosti, prípadne nevhodnosť žiadateľa pre vytvorenie účtu

5. Administrátor skontroluje informácie v žiadosti a klikne na tlačidlo - "Zamietnuť žiadosť"
6. Systém zobrazí textové pole, kde je potrebné napísať dôvod zamietnutia žiadosti
7. Administrátor vyplní dôvod zamietnutia a následne klikne na tlačidlo - "Zamietnuť žiadosť" (alebo Potvrdiť... prípadne niečo podobné)
8. Systém odošle správu o zamietnutí spolu s dôvodom na zadaný email v žiadosti
9. Systém označí žiadosť v databáze za zamietnutú
10. Systém presmeruje používateľa do zoznamu všetkých nevybavených žiadostí o registráciu

UC03 – Prihlásenie používateľa

Ako farmár s vytvoreným kontom sa chcem vedieť prihlásiť a dostať sa k funkcionalitám systému, ktoré sú sprístupnené len po prihlásení do systému.

Predpoklad - farmár sa nachádza na home page webového rozhrania systému.

1. Farmár klikne na tlačidlo - "Prihlásiť sa"
2. Systém zobrazí okno s možnosťou vyplnenia prihlasovacích údajov
3. Farmár vyplní prihlasovacie údaje a klikne na tlačidlo - "Prihlásiť sa"
4. Systém skontroluje údaje
5. Systém presmeruje používateľa na úvodnú stránku rozhrania prihláseného používateľa

Alternatívny scenár (od kroku 5) - nekorektné prihlasovacie údaje

5. Systém vypíše upozornenie, že prihlasovacie údaje nie sú správne

UC04 – Pridanie poľa

Ako prihlásený farmár chcem pridať pole, na ktorom pestujem technické konope do systému, aby som k ho mohol priradovať k jednotlivým nahratým dátam z merania.

Predpoklad - prihlásený farmár sa nachádza na úvodnej stránke po prihlásení a má fyzické pole, na ktorom chce vykonávať merania vzoriek a ešte ho nemá pridané v systéme

1. Prihlásený farmár v menu zvolí položku - "Moje polia"
2. Systém zobrazí zoznam jeho polí
3. Prihlásený farmár klikne na tlačidlo - "Pridať pole"
4. Systém zobrazí formulár, kde je možné vyplniť informácie o pridávanom poli (názov, opis, geolokácia..)
5. Prihlásený farmár vyplní potrebné informácie
6. Prihlásený farmár klikne na tlačidlo - "Pridať pole"
7. Systém pridá položku do zoznamu polí a presmeruje používateľa na zoznam jeho polí

UC05 – Pridanie druhu rastliny

Ako administrátor chcem pridať druh rastliny technického konope, aby ju mohli prihlásení farmári priradovať jednotlivým nahratým dátam z merania, prípadne, aby som ju mohol ja priradiť prihláseným farmárom pri nahrávaní údajov zo spektrometra.

Predpoklad - administrátor sa nachádza na úvodnej stránke po prihlásení a má v úmysle pridávať druh rastliny, ktorý ešte nefiguruje v systéme

1. Administrátor v menu zvolí položku - "Druhy rastlín"
2. Systém zobrazí zoznam všetkých druhov rastlín, ktoré sú v systéme uložené
3. Administrátor vyplní políčko "Druh konopy"
4. Administrátor klikne na tlačidlo "Pridať druh"
5. Systém pridá vyplnený druh rastliny do zoznamu druhov rastlín a
6. Systém nanovo načíta aktuálnu stránku s novým druhom v tabuľke

UC06 – Import nameraných dát

Ako používateľ chcem importovať namerané dáta zo spektrometra, aby ich bolo možné vyhodnotiť ML modelom.

Predpoklad - prihlásenie používateľa do systému, používateľ sa nachádza na úvodnej obrazovke po prihlásení, existencia súboru s nameranými dátami

1. Používateľ klikne na tlačidlo - "Nové meranie"
2. Systém zobrazí formulár s možnosťou nahratia súboru
3. Používateľ nahrá súbor s dátami do systému
4. Systém skontroluje formát súboru
5. Používateľ vyplní potrebné údaje vo formulári

6. Používateľ klikne na tlačidlo - "Pridať meranie"
7. Systém vytvorí novú entitu meranie, ktorú pridá do zoznamu meraní pre používateľa
8. Systém nahrá a uloží súbor na serveri
9. Systém presmeruje používateľa na inú obrazovku

UC07 – Import nameraných dát prihláseným farmárom

Ako prihlásený farmár chcem importovať namerané dáta to spektrometra, aby ich bolo možné vyhodnotiť ML modelom.

Predpoklad - prihlásenie farmára do systému, farmár sa nachádza na úvodnej obrazovke po prihlásení, existencia súboru s nameranými dátami

1. Prihlásený farmár klikne na tlačidlo - "Nové meranie"
2. Systém zobrazí formulár s možnosťou nahrať súboru, výberu druhu rastliny a poľa podľa dostupných u farmára a ďalšími možnými údajmi o meraní
3. Prihlásený farmár nahrá súbor s dátami do systému
4. Systém skontroluje formát súboru
5. Prihlásený farmár vyplní potrebné údaje vo formulári - dátum odobratia vzorky, poznámky k meraniu, vzorke, vyberie druh rastliny
6. Prihlásený farmár klikne na tlačidlo - "Pridať meranie"
7. Systém vytvorí novú entitu meranie, ktorú pridá do zoznamu meraní pre používateľa
8. Systém nahrá a uloží súbor na serveri
9. Systém presmeruje prihláseného farmára na obrazovku s možnosťou spustenia modelu nad pridaným súborom
10. Prihlásený farmár klikne na tlačidlo - "Spustiť model"
11. Systém presmeruje používateľa na informačnú stránku o spustení modelu
12. Systém prepočíta výsledky k danému vstupnému súboru a priradí ich k meraniu

Alternatívny scenár (od kroku 5) - farmár chce pridať nový druh rastliny

Predpoklad - prihlásený farmár chce k meraniu pridať druh rastliny, ktorý ešte nie je v systéme zaevidovaný

5. Prihlásený farmár vyplní potrebné údaje vo formulári - dátum odobratia vzorky, poznámky k meraniu, vzorke a zadá nový druh rastliny
6. Prihlásený farmár klikne na tlačidlo - "Pridať meranie"
7. Systém vytvorí novú entitu druhu rastliny, ktorú pridá do databázy
8. Systém vytvorí novú entitu meranie, ktorú pridá do zoznamu meraní pre používateľa
9. Systém presmeruje prihláseného farmára na obrazovku s možnosťou spustenia modelu nad pridaným súborom
10. Prihlásený farmár klikne na tlačidlo - "Spustiť model"
11. Systém presmeruje používateľa na informačnú stránku o spustení modelu
12. Systém prepočíta výsledky k danému vstupnému súboru a priradí ich k meraniu

UC08 – Import nameraných dát administrátorom

Ako administrátor chcem importovať namerané dáta spektrometrom o vzorkách, ktoré zaslal farmár na vyhodnotenie, aby ich pre farmára bolo možné vyhodnotiť ML modelom.

Predpoklad - prihlásenie administrátora do systému, administrátor sa nachádza na úvodnej obrazovke po prihlásení, existencia súboru s nameranými dátami, ktoré boli vyhotovené na základe vzorky, ktorú farmár zaslal na spoločné centrálné miesto, existencia identifikácie farmára a jeho vzorky

1. Administrátor klikne na tlačidlo -“Nové meranie”
2. Systém zobrazí formulár s možnosťou nahratia súboru ,výberu používateľa z dropdownu, výberu druhu rastliny a zadania dátumu
3. Administrátor nahrá súbor s dátami do systému
4. Systém skontroluje formát súboru
5. Administrátor vyplní potrebné údaje vo formulári - vyberie farmára, vyberie druh rastliny a zadá dátum zberu vzorky
6. Administrátor klikne na tlačidlo - “Pridať meranie”
7. Systém vytvorí novú entitu meranie, ktorú pridá do zoznamu meraní pre príslušného používateľa
8. Systém nahrá a uloží súbor na serveri
9. Systém presmeruje administrátora na úvodnú obrazovku

UC09 – Editácia merania

Ako prihlásený farmár chcem mať možnosť editovať meranie, aby som mohol doplniť, prípadne opraviť údaje o meraní.

Predpoklad - prihlásenie farmára do systému, farmár sa nachádza na úvodnej obrazovke po prihlásení, existencia aspoň jedného merania priradeného k jeho účtu

1. Prihlásený farmár klikne na tlačidlo - “Moje merania”
2. Systém zobrazí zoznam meraní používateľa
3. Prihlásený farmár nájde a zvolí meranie, ktoré chce editovať
4. Systém zobrazí detail merania
5. Prihlásený farmár klikne na tlačidlo “Editovať”
6. Systém otvorí predvyplnený formulár podľa aktuálnych údajov pre zvolené meranie
7. Prihlásený farmár edituje údaje, podľa vlastnej potreby
8. Prihlásený farmár klikne na tlačidlo “Uložiť”
9. Systém uloží aktuálny stav údajov a presmeruje prihláseného farmára na zoznam jeho meraní

UC10 - Spustenie vyhodnotenia merania

Ako prihlásený farmár chcem vedieť spustiť model na vyhodnotenie merania, ktoré ešte nebolo vyhodnotené, aby som mohol získať výsledky z tohto merania.

Predpoklad - prihlásenie farmára do systému, farmár sa nachádza na úvodnej obrazovke po prihlásení, existencia aspoň jedného nahratého merania priradeného k jeho účtu, ktoré ešte nie je vyhodnotené

1. Prihlásený farmár klikne na tlačidlo - "Moje merania"
2. Systém zobrazí zoznam meraní používateľa
3. Prihlásený farmár nájde meranie, ktoré chce vyhodnotiť v tabuľke a dvojklikom ho otvorí
4. Systém zobrazí obrazovku pridanie nového merania v 2. kroku - spustenie modelu
5. Prihlásený farmár klikne na tlačidlo - "Spustiť model"
6. Systém presmeruje používateľa na informačnú stránku o spustení modelu
7. Systém prepočíta výsledky k danému vstupnému súboru a priradí ich k meraniu

UC11 – Otvorenie vizualizácie výsledkov

Ako prihlásený farmár chcem vizualizovať výsledky merania, aby som mohol vidieť stav látok v danej vzorke.

Predpoklad - prihlásenie farmára do systému, farmár je na úvodnej obrazovke po prihlásení, existencia aspoň jedného merania s výsledkami z modelu, priradeného k jeho účtu

1. Prihlásený farmár klikne na tlačidlo "Moje merania"
2. Systém zobrazí zoznam meraní používateľa
3. Prihlásený farmár nájde a dvojklikom zvolí meranie v zozname meraní.
4. Systém otvorí vizualizáciu výsledkov merania vzorky

UC12 – Otvorenie porovnania vizualizácií výsledkov 2 vzoriek

Ako prihlásený farmár chcem vizualizovať porovnanie 2 výsledkov nameraných vzoriek, aby som bol schopný jednoducho porovnať a nájsť rozdiely medzi výsledkami daných 2 meraní.

Predpoklad - prihlásenie farmára do systému, farmár sa nachádza na úvodnej obrazovke po prihlásení, existencia aspoň 2 meraní priradených k účtu farmára, ktoré má už poskytnuté výsledky z ML modelu

1. Include UC11 - Otvorenie vizualizácie výsledkov
2. Prihlásený farmár klikne na tlačidlo "Porovnať merania"
3. Systém otvorí okno, kde je možné vybrať meranie na porovnanie zo zoznam všetkých meraní
4. Prihlásený farmár vyberie meranie pre porovnanie s aktuálne otvoreným
5. Systém zobrazí porovnanie obidvoch meraní vedľa seba

UC13 – Odstránenie merania

Ako prihlásený farmár chcem mať možnosť odstrániť meranie, aby som mohol sprehľadniť moje merania, prípadne odstrániť nekorektné meranie.

Predpoklad - prihlásenie farmára do systému, farmár sa nachádza na úvodnej obrazovke po prihlásení, existencia aspoň jedného merania priradeného k jeho účtu

1. Prihlásený farmár klikne na tlačidlo - "Moje merania"
2. Systém zobrazí zoznam meraní používateľa
3. Prihlásený farmár nájde meranie, ktoré chce odstrániť a klikne na tlačidlo v rámci záznamu - "Odstrániť meranie" (príp. "X")
4. Systém zobrazí okno s potvrdením akcie ("Naozaj chcete...")
5. Prihlásený farmár klikne na tlačidlo "Áno"
6. Systém odstráni meranie + všetky jeho dáta, vymaže meranie zo zoznamu meraní a zatvorí okno s potvrdením

Alternatívny scenár (od kroku 5) - farmár nechce meranie odstrániť

Predpoklad - prihlásený farmár si rozmyslel, resp. nechce vymazať meranie

5. Prihlásený farmár klikne na tlačidlo "Nie"
6. Systém zatvorí okno s potvrdením

UC14 – Deaktivácia účtu

Ako administrátor chcem vedieť deaktivovať používateľské konto v prípade potreby, aby som mohol zabrániť zneužitiu, zamedzil neoprávnený prístup(ak už používateľ nemá právo používať - napr. už nie je pestovateľ alebo už nespĺňa podmienky schválenia registrácie), zamedzil prístup používateľovi, ktorý zneužil portál, prípadne odstránil nepoužívaný účet.

Predpokladom je prihlásenie v administrátorskom rozhraní - na úvodnej stránke tohto rozhrania a existujúci aspoň jeden účet farmára.

1. Administrátor klikne na tlačidlo - "Používatelia"
2. Systém zobrazí zoznam všetkých používateľov
3. Administrátor klikne na jedného z používateľov - zvolí používateľa
4. Systém zobrazí detail používateľa spolu so všetkými jeho informáciami
5. Administrátor klikne na tlačidlo - "Deaktivácia účtu"
6. Systém označí účet v databáze za deaktivovaný - už nebude možné prihlásenie, ponechá si ale všetky jeho údaje a merania
7. Systém presmeruje administrátora späť na zoznam používateľov

UC15 – Vymazanie druhu rastliny

Ako administrátor chcem vymazať druh rastliny technického konope, aby sa už ďalej tento druh v systéme nemohol používať.

Predpoklad - administrátor sa nachádza na úvodnej stránke po prihlásení a má v úmysle vymazať druh rastliny, ktorý sa nachádza v zozname druhov a zároveň nie je priradené ku žiadnemu meraniu

1. Administrátor v menu zvolí položku - "Druhy rastlín"
2. Systém zobrazí zoznam všetkých druhov rastlín, ktoré sú v systéme uložené
3. Administrátor nájde v zozname požadovaný druh rastliny a klikne na tlačidlo "X" pri danej položke
4. Systém zobrazí modálne okno s potvrdením
5. Administrátor potvrdí vymazanie

Systém odstráni zvolený druh rastliny zo zoznamu druhov rastlín a nanovo načíta aktuálnu stránku bez vymazaného druhu v tabuľke

Diagram UC - Správa používateľov

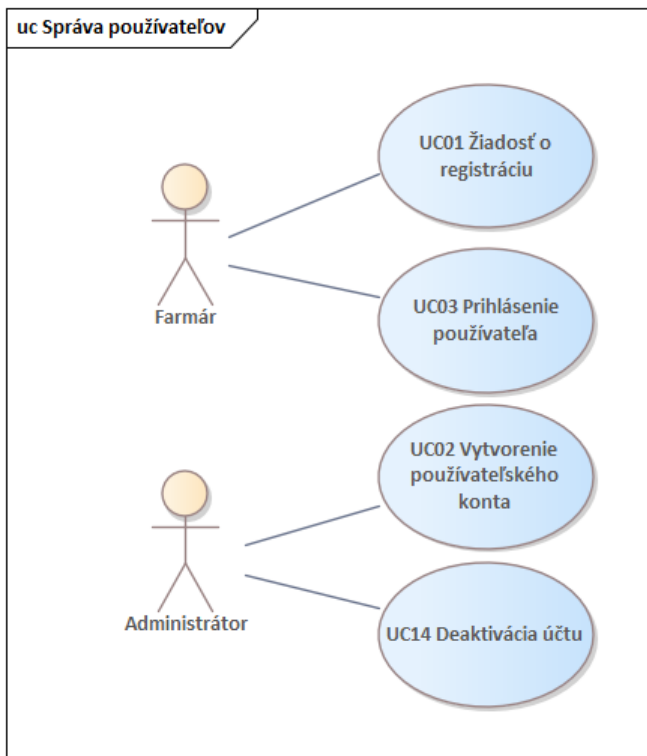


Diagram UC - Vkladanie a mazanie údajov

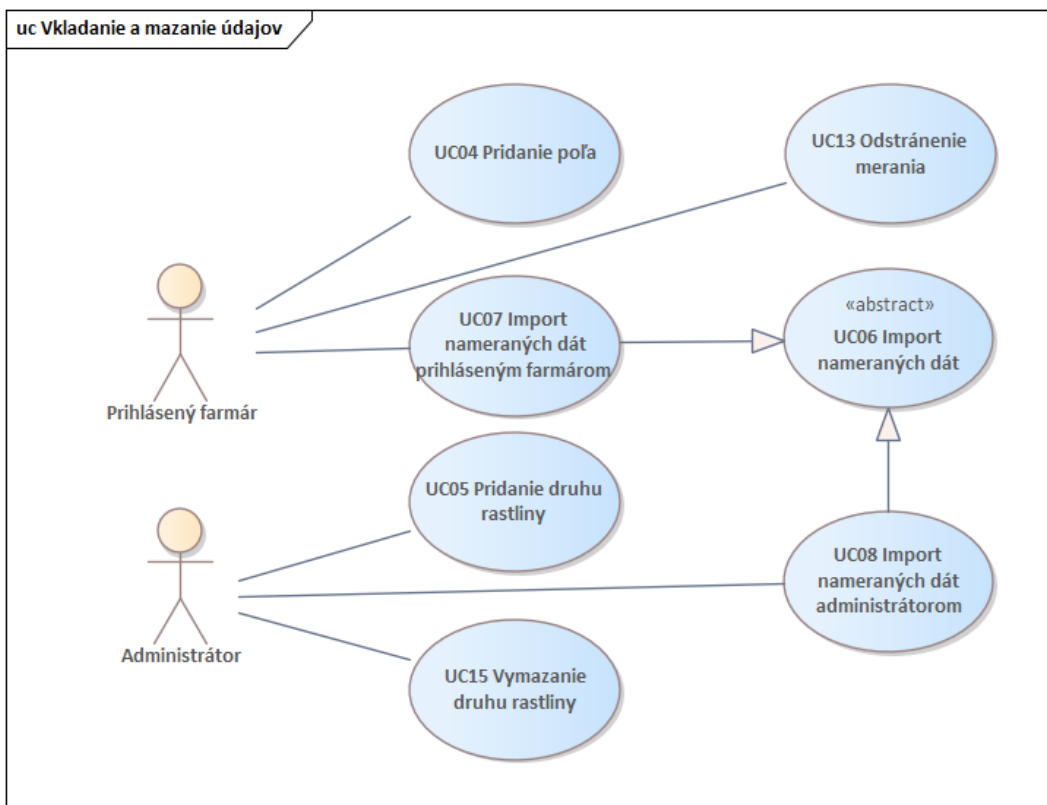
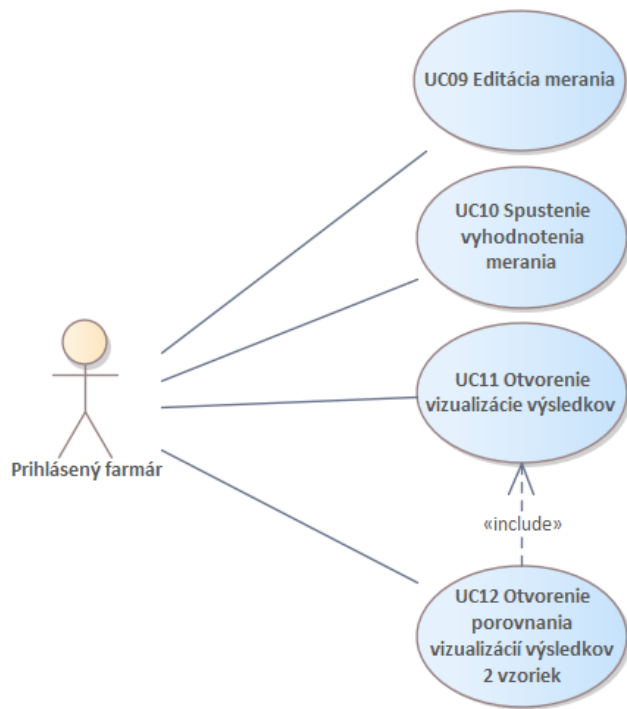


Diagram UC - Práca s meraniami

uc Práca s meraniami



4.2. Návrh webovej aplikácie

Pri návrhu webovej aplikácie pokračujeme vyhotovením konkrétnych návrhov používateľského rozhrania, ktoré majú slúžiť ako podklad pri implementácii. Návrhy sme vypracovali v online nástroji Figma. Návrh je rozdelený na 3 súvislé celky: homepage, administrátorské rozhranie a používateľské rozhranie. Pri každom celku sú zobrazené obrazovky a slovné popisy k nim. Zobrazené návrhy sú výsledkom druhej iterácie, pri ktorej sme sa snažili vylepšiť koncepty z pôvodnej, prvej iterácie navrhovania.

4.2.1. Návrh homepage

Úlohou homepage stránky je, aby zaujala na prvý pohľad, čo sme sa snažili dosiahnuť dizajnom, ktorý je zobrazený na nasledujúcom obrázku.



Zvyšné obrázky homepage predstavujú pokračovanie stránky po scrollovaní smerom dole. Táto časť má informačný charakter, kde sa môže potenciálny používateľ dočítať bližšie a podrobnejšie informácie o produkte, prípadne o iónovom spektrometri a pod. Na záver homepage je umiestnený registračný formulár, pomocou ktorého môže používateľ podať žiadosť o registráciu.

Na našom potráli si môžete vkladať namerané vzorky z Vašej technickej konopy a naše modely strojového učenia Vám pomôžu určiť deň, kedy zbierať odrodu tak, aby dosahovala najlepšie hodnoty CBD. Takto Vám dokážeme pomôcť pri zabezpečení kvalitných výrobkov z Vašej konopy.



Ako to funguje?



Naše modely sú natréňované zo vzoriek technickej konopy vypestovanej na Slovensku. V modeli sú vzorky rôznych druhov konope, čo zabezpečuje jeho väčšiu všeobecnosť. Naše vzorky sú obsahujú informáciu o ich zrelosti. Aby sme vedeli odporučiť poľnohospodárom, kedy má zbierať, disponujeme aj vzorkami, ktoré sú zrelé na zber. Vďaka týmto vzorkám dokáže náš model určiť, ako ďaleko sa Vaša nameraná vzorka nachádza od ideálneho stavu.

Meranie iónovým pohyblivostným spektrometrom

Meranie vzoriek z nášho modelu bolo pokryté iónovým pohyblivostným spektrometrom od firmy MaSaTECH. IMS funguje na princípe zohrievania meranej vzorky a uvoľňovania jednotlivých látok, ktoré sú zachytávané kapilárami. Pri každom meraní vzniklo 800 spektier, ktoré po vykreslení do 2D mapy predstavujú postupné uvoľňovanie látok zo vzorky. Po analýze problému sme boli schopní určiť látku CBD, ktorá sa nachádzala v spektrách.

MaSaTECH



Máte záujem o registráciu? Môžete vyplniť formulár:

Meno a priezvisko

Email

Telefónne číslo

Názov firmy

IČO

Poslať žiadosť

4.2.2. Návrh administrátorského rozhrania

Administrátorské rozhranie bude slúžiť na riadenie používateľov, správu žiadostí o registráciu a prípadné importovanie dát zo spektrometra. Import dát administrátorom je možné v prípade kedy samotný používateľ nebude mať k dispozícii spektrometer a zašle svoju vzorku na miesto, kde budú namerané jej hodnoty a následne importované do systému a pridelené používateľovi.

Úvodná stránka administrátora po prihlásení bude zobrazovať základnú štatistiku o používateľoch a nahratých vzorkách.



V sekcii používateľa bude administrátor môcť vidieť zoznam všetkých používateľov s možnosťou filtrácie a otvorenia detailu po kliknutí na konkrétneho používateľa.

HempTeam Prihlásený: Admin | Odhlásiť sa

Zoznam používateľov

Meno používateľa: Email:
 Tel. č.: **Vyhľadať**

Meno	Email
Juraj Pokorný	juraj.pokorny@gmail.com
Diana Mrkvíčková	diana.mrkvicikova@gmail.com
Dávid Čierny	david.ciorny@gmail.com
Jožko Mrkvíčka	jozko.mkrvicicka@gmail.com
Mária Danišová	maria.danisova@gmail.com
Zdeno Horný	zdeno.horny@gmail.com
Ján Dolný	jan.dolny@gmail.com
Jana Mrkvíčková	jana.mrkvicikova@gmail.com

1. - 8. z 15 používateľov < 1 2 3 >

Copyright © 2021 - HempTeam FIIT STU

V rámci detailu budú zobrazené základné informácie o používateľovi. Pri detaile bude v prípade potreby možné deaktivovať účet používateľa.

HempTeam Prihlásený: Admin | Odhlásiť sa

Detail používateľa

Meno: Jozef Mrkvíčka
 Email: jozko.mrkvicicka@gmail.com
 Telefón: +421 900 000 000
 IČO: 12345678

Posledná aktivita: 10. 10. 2010
 Počet nahratých vzoriek: 24
 Dátum nahratia poslednej vzorky: 10. 10. 2010

Deaktivovať účet

Copyright © 2021 - HempTeam FIIT STU

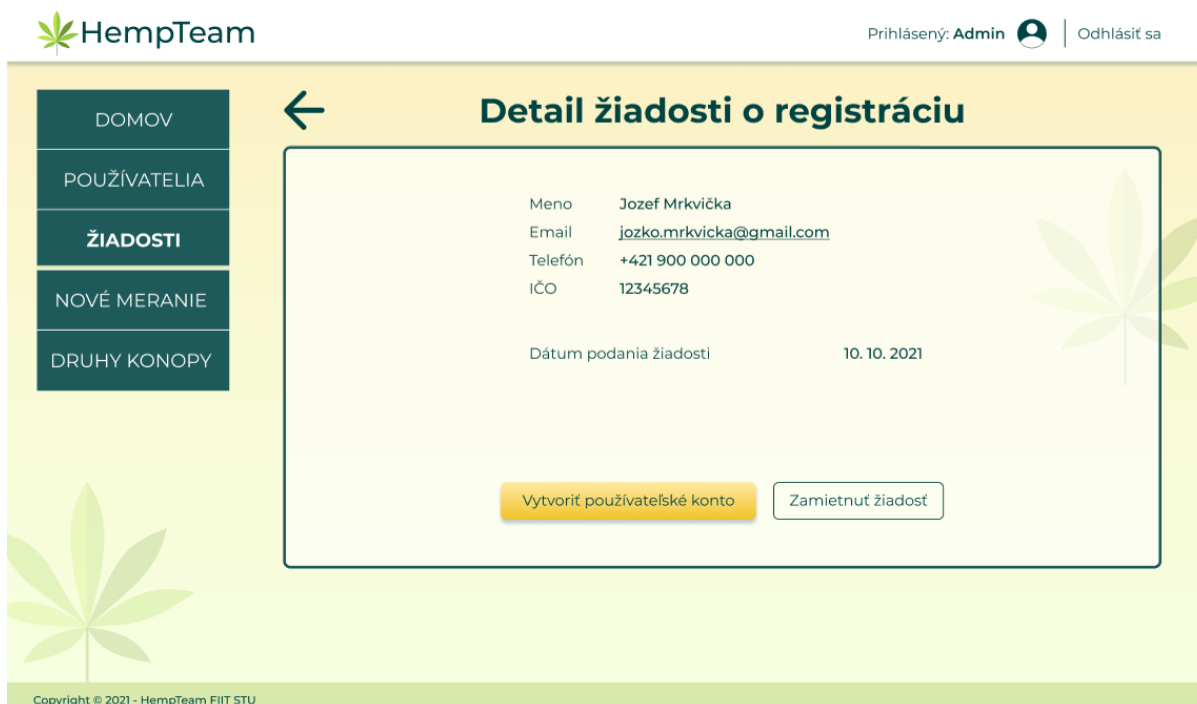
V sekcii žiadosti o registráciu bude mať administrátor k dispozícii všetky podané a nevybavené žiadosti o registráciu v zozname. Po kliknutí na konkrétnu žiadosť sa otvorí detail tejto žiadosti.



Meno	Dátum podania žiadosti
Juraj Pokorný	24.11.2021
Diana Mrkvičková	19.3.2020
Dávid Čierny	12.8.2021
Jožko Mrkvička	15.9.2021
Mária Danišová	18.6.2020
Zdeno Horný	19.7.2020
Ján Dolný	25.4.2021
Jana Mrkvičková	18.10.2021

1. - 8. z 15 žiadostí

V detaile žiadosti budú zobrazené základné informácie, ktoré poskytol používateľ pri podaní žiadosti. Na tejto obrazovke bude možné zamietnuť žiadosť alebo vytvoriť na základe danej žiadosti nové používateľské konto.



Meno: Jozef Mrkvička
Email: jozko.mrkvicka@gmail.com
Telefón: +421 900 000 000
IČO: 12345678

Dátum podania žiadosti: 10. 10. 2021

Vytvoriť používateľské konto Zamietnuť žiadosť

V sekcií pridanie nového merania bude administrátorovi umožnené importovať súbor s dátami zo spektrometra. Okrem nahratia súboru je potrebné priradiť používateľa, druh

konopy a dátum zberu vzorky. Tieto údaje budú potrebné pre používateľa pre jednoznačnú identifikáciu merania.

 Prihlásený: Admin  | Odhlásiť sa

Pridanie nového merania

DOMOV
POUŽÍVATELIA
ŽIADOSTI
NOVÉ MERANIE
DRUHY KONOPY

Meno používateľa

Druh konopy

Dátum zberu vzorky

Potiahnite alebo vyberte súbory



Vybrať súbory

Pridať meranie

Copyright © 2021 - HempTeam FIIT STU

Administrátor bude mať k dispozícii editor na pridávanie a mazanie jednotlivých druhov konopy. Pomocou zadania reťazca a stlačenia tlačidla "Pridať druh" je možné druh konopy pridať. Každý druh je možné zmazať pomocou tlačidla "X".

 Prihlásený: Admin  | Odhlásiť sa

Zoznam druhov konopy

DOMOV
POUŽÍVATELIA
ŽIADOSTI
NOVÉ MERANIE
DRUHY KONOPY

Druh konopy Pridať druh

Druh konopy	Počet meraní druhu	
Konope siate	5	X
Konope indické	46	X
Konope rumištné	9	X
Iný druh konope 123	134	X
Ďalší druh	58	X
Opäť iný druh	5	X
Nasledujúci druh	10	X
A ďalší iný druh	12	X
Predposledný druh	8	X
Posledný druh	0	X


1. - 10. z 28 druhov rastlín < 1 2 3 >

Copyright © 2021 - HempTeam FIIT STU

4.2.3. Návrh používateľského rozhrania

V tejto sekcii opíšeme návrh webového prostredia pre našich používateľov (pestovateľov konope). V tejto časti má používateľ možnosť pridávať merania, spravovať ich, spravovať svoje polia a prezerať výsledky modelov z nameraných dát, ktoré boli získané nameraním konkrétnych vzoriek konopy pomocou spektrometra.

Po kliknutí na tlačidlo “Prihlásiť sa” v menu na homepage je používateľ presmerovaný na stránku prihlásenia. Ak nie je ešte registrovaný, môže kliknúť na tlačidlo “Požiadajte o registráciu” ktoré ho presmeruje na registračný formulár.

HempTeam Prihlásiť sa  Registrovať sa

Prihlásenie do portálu

Email

Heslo

[Prihlásiť sa](#)

Ak ešte nemáte vytvorené konto, môžete požiadať o registráciu:

[Požiadajte o registráciu](#)

HempTeam

Tento projekt vznikol ako výsledok tímového projektu na Fakulte informatiky a informačných technológií.

MaSaTECH


Meranie vzoriek technickej konopy bolo zabezpečené product ownerom Martinom Sabom, Mgr., technológiami firmy MaSaTECH.


HempBona

Vzorky technickej konopy boli darované firmou HempBona s poliami na východe Slovenska pri obci Pčelín.



Kontakt

Fakulta informatiky a informačných technológií
STU v Bratislave

 Ilkovičova 2, 842 16 Bratislava

 tp.tim2@gmail.com

Žiadosť o registráciu slúži pre nového užívateľa na registráciu v ktorej musí vyplniť základné údaje potrebné pre zaregistrovanie.

 Prihlásiť sa  | Registrovať sa

Máte záujem o registráciu? Môžete vyplniť formulár:

Meno a priezvisko


Email

Telefónne číslo



Názov firmy

IČO

[Poslať žiadosť](#)



Po prihlásení sa užívateľovi zobrazí prehľad jeho jednotlivých meraní. V rámci meraní je možné ich aj vyhľadávať. Každé meranie je možné v prípade potreby zmazať pomocou tlačidla "X".

 Prihlásený: **Jožko**  | Odhlásiť sa

Moje merania

[DOMOV](#)
[**MERANIA**](#)
[NOVÉ MERANIE](#)
[POLIA](#)
[MOJE ÚDAJE](#)

Názov poľa Dátum zberu vzorky

Druh konopy

[Vyhľadať](#)

Druh konopy	Názov poľa	Dátum zberu vzorky	
Cannakomp	Juhovýchodné pole	17. 6. 2021	X
Cannakomp	Juhovýchodné pole	17. 6. 2021	X
Denise	Južné pole	12. 6. 2021	X
Fibranova	Južné pole	7. 6. 2021	X
Fibranova	Južné pole	1. 6. 2021	X

1. - 5. z 5 meraní [<](#) [1](#) [>](#)

Copyright © 2021 - HempTeam FIIT STU

V časti pridanie nového merania môže používateľ nahráť nové meranie medzi svoje merania. Táto činnosť prebieha podľa “stepera” v hornej časti v troch krokoch. V prvom kroku používateľ nahrá súbor a vyplní základné údaje.

 Prihlásený: Jožko  | [Odhlásiť sa](#)

Pridanie nového merania

1 — 2 — 3
Pridanie súborov z merania — Spustenie modelu — Počítanie výsledkov

Druh konopy

Dátum zberu vzorky

Názov poľa

Potiahnite alebo vyberte súbory



Copyright © 2021 - HempTeam FIIT STU

V druhom kroku je už úspešne pridané meranie, ktoré ešte nemá vypočítané výsledky pomocou modelu. Tu je možné model spustiť.

 Prihlásený: Jožko  | [Odhlásiť sa](#)

Pridanie nového merania

1 — 2 — 3
Pridanie súborov z merania — Spustenie modelu — Počítanie výsledkov



Nahrávanie súborov prebehlo úspešne pre nasledovnú vzorku:

Druh konopy

Dátum zberu vzorky

Názov poľa

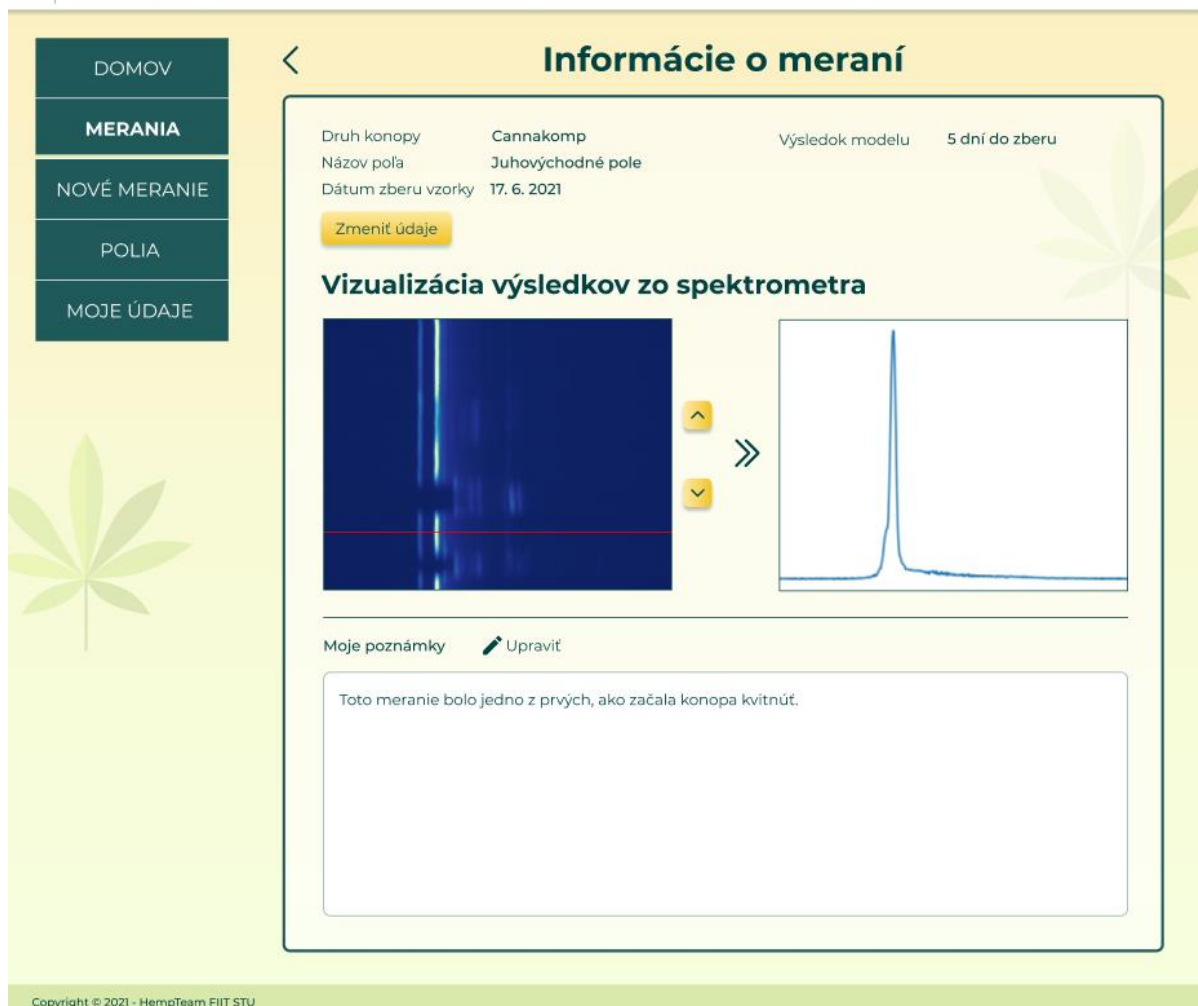
Na tejto vzorke teraz môžete spustiť model kliknutím na tlačidlo nižšie:

Copyright © 2021 - HempTeam FIIT STU

V treťom kroku je možné vidieť, že je pridané merania a už sa aj spustil model na vypočítanie výsledkov. Čaká sa len na ich vypočítanie.

The screenshot displays the HempTeam web interface. At the top left is the HempTeam logo. At the top right, it shows the user is logged in as 'Jožko' with a profile icon and an 'Odhlásiť sa' (Log out) link. A dark green sidebar on the left contains navigation buttons: 'DOMOV', 'MERANIA', 'NOVÉ MERANIE' (highlighted), 'POLIA', and 'MOJE ÚDAJE'. The main content area is titled 'Pridanie nového merania' (Adding a new measurement). It features a progress indicator with three steps: 1. 'Pridanie súborov z merania' (Adding files from measurement), 2. 'Spustenie modelu' (Running the model), and 3. 'Počítanie výsledkov' (Calculating results). Step 2 is currently active, indicated by a green checkmark in a box. Below this, a message states: 'Spustenie modelu prebehlo úspešne. Výsledky si budete môcť pozrieť už o malú chvíľu kliknutím na nasledujúce tlačidlo:' (Model running was successful. You will be able to view the results in a moment by clicking the following button:). A yellow button labeled 'Zobraziť detail' (View details) is positioned below the message. The background of the main area features a faint green hemp leaf graphic. At the bottom left, a footer reads 'Copyright © 2021 - HempTeam FIIT STU'.

Po ukončení počítania modelu, prípadne vždy po otvorení detailu merania v zozname meraní je možné si prezerať informácie o meraní. V prípade potreby je tu umožnené editovať údaje o meraní.



Informácie o meraní

Druh konopy: Cannakomp Výsledok modelu: 5 dní do zberu
Názov poľa: Juhovýchodné pole
Dátum zberu vzorky: 17. 6. 2021

[Zmeniť údaje](#)



Vizualizácia výsledkov zo spektrometra

Moje poznámky [Upraviť](#)

Toto meranie bolo jedno z prvých, ako začala konopa kvitnúť.

Copyright © 2021 - HempTeam FIIT STU

V sekcii polia môže používateľ spravovať svoje polia. Po kliknutí na túto sekciu v menu sa mu zobrazí zoznam jeho existujúcich polí so základnými informáciami.

 Prihlásený: **Jožko**  | [Odhlásiť sa](#)

Moje polia



Názov poľa [Vyhľadať](#) [Pridať nové pole](#)

Názov poľa	Súradnice	Počet vzoriek
Juhovýchodné pole	48.961845, 21.718259	2
Južné pole	48.961845, 21.718259	3
Nové pole	48.961845, 21.718259	0
Južná časť poľa	48.961845, 21.718259	0
Na čistinke	48.961845, 21.718259	0

1. - 5. z 5 polí [<](#) [1](#) [>](#)

Copyright © 2021 - HempTeam FIIT STU

Po kliknutí na konkrétne pole sa užívateľovi zobrazia podrobné informácie o poli. Tu je umožnená aj editácia informácií o poli.

 Prihlásený: **Jožko**  | [Odhlásiť sa](#)


Upraviť údaje: Juhovýchodné pole

Názov poľa

Súradnica 1 (zemepisná šírka)

Súradnica 2 (zemepisná výška)



Ukážka polohy na mape:



[Upraviť údaje](#)

Copyright © 2021 - HempTeam FIIT STU

Pri zozname všetkých polí sa nachádza tlačidlo "Pridať nové pole". Po jeho stlačení sa používateľovi zobrazí obrazovka Pridať nové polu. Na nej je umožnené používateľovi pridávať polia do zoznamu jeho polí.

 Prihlásený: **Jožko**  | [Odhlásiť sa](#)

- DOMOV
- MERANIA
- NOVÉ MERANIE
- POLIA**
- MOJE ÚDAJE


Pridať nové pole

Názov poľa

Súradnica 1 (zemepisná šírka)

Súradnica 2 (zemepisná výška)

Ukážka polohy na mape:



[Pridať nové pole](#)

Copyright © 2021 - HempTeam FIIT STU

4.3. Implementácia webovej aplikácie

Webovú aplikáciu implementuje vo frameworku Laravel. Používame pri to programovací jazyk PHP. Pre tvorbu frontendu používame HTML, CSS a JS. HTML je generované pomocou frameworku Laravel. Používame pri tom Blade. Pre tvorbu CSS používame less. Kompilujeme .less súbory pomocou Gulp-u. Vzniká nám jeden style.css súbor, kde sa nachádzajú všetky štýly. Aplikáciu máme pripojenú na mysql databázu, ktorá nám beží na serveri.

Pre aplikáciu máme vytvorený spoločný git repozitár na GitHube. Nachádzajú sa tam všetky zdrojové kódy. Na základe aktuálnej verzie tohto repozitára máme nastavený deploy na server. Po ukončení určitej verzie aplikácie je spravená inštalácia na server. Robíme to pravidelne po dokončení každého šprintu.

V aktuálnom stave projektu je implementovaná logika aplikácie, ktorá sa týka správy používateľov. Konkrétne sa jedná o používateľské scenáre UC1 - UC3. Spravili sme frontend, ktorý odzrkadľuje návrhy obrazoviek. Na základe vytvoreného frontendu sme spravili backend logiku. Jednalo sa o vytvorenie tabuliek v databáze. Následne sme týmto tabuľkám spravili model vo frameworku Laravel. Na základe vytvoreného modelu sme spravili oprávnnenia, ktoré sú jadrom celej logiky modelu používateľov. Je to potrebné hlavne kvôli zabezpečeniu a ošetrovaniu výnimočných stavov. Následne sme prepojili dáta z databázy s frontendom.

Okrem toho sme implementovali základné pridanie merania pre používateľov, kedy používateľ môže zatiaľ bez pridania súboru vytvoriť entitu merania. Takisto môže vidieť všetky svoje vytvorené merania v zozname meraní. Zoznam meraní používateľa je zobrazený po jeho prihlásení.

4.4. Testovanie webovej aplikácie

Testy webovej aplikácie budú realizované pomocou Robot framework, čo je nadstavba nad Selenium library. Budeme na to využívať unit testy (v preklade jednotkové testy), čo znamená, že budú overovať funkčnosť webovej aplikácie a budú overovať, či sa počas vývoja aplikácie nezaviedla nová chyba už do fungujúceho systému. Testovacie scenáre budú realizované pre overenie veškerej funkcionality webovej aplikácie.

4.5. Analýza dátového modelu

4.5.1. Úvod do problematiky

Cieľom poľnohospodárov je, aby zbierali technickú konopu vtedy, keď má najvyšší obsah CBD. Na to, aby bol zistený aktuálny obsah kanabinooidov, medzi ktoré patrí aj CBD, je potrebné zaslať vzorku do laboratória. V našej práci používame na zber dát iónový spektrometer, ktorým meriame vzorky technickej konopy nazbierané v rôznych etapách dozrievania.

Vytvorením modelu strojového učenia chceme pomôcť poľnohospodárom na základe výsledkov ich vzoriek určiť čas, kedy zbierať konopu tak, aby výrobky z nej dosahovali čo najvyššiu kvalitu.

Hlavnou motiváciou pre vypracovanie projektu v tejto téme je okrem aktuálnej popularity, ktorú CBD na Slovensku naberá, aj potenciálny prospech výskumu CBD. Chceme využiť naše doposiaľ nadobudnuté skúsenosti a vedomosti s cieľom podporovania vedy a výskumu v oblastiach, ktorým má zmysel sa venovať. V téme vidíme priestor pracovať so zaujímavou sadou dát a možnosť nachádzať v nich mnoho zaujímavých vzorov, identifikovať ich zaujímavé črty a v neposlednom rade verifikovať naše domnienky, ktoré vychádzajú zo skúmania dát.

4.5.2. Súvisiace práce

V dnešnej dobe sa iónová pohyblivostná spektrometria využíva na rozpoznanie výbušnín v neznámej látke, zistenie pravosti niektorých potravín, identifikácia čerstvosti mäsa a mnoho ďalšieho. Množstvo rôznych prác často používa množstvo rôznych prístupov spracovania dát a našou úlohou bolo zjednotiť si v tíme prístup, akým budeme dáta reprezentovať, spracovávať a následne na základe nich trénovať model.

V práci od Garrido-Delgado, R. a kol. [2] autori pomocou iónového pohyblivostného spektrometra klasifikovali rôzne druhy vína. Využitie je najmä pri odhaľovaní falšovania vína (či víno neobsahuje iné látky ako deklaruje výrobca), pri rozpoznaní kvality vína (s cieľom niektorých vinárov expandovať na svetový trh) alebo pri garantovaní pôvodu vína. V práci skúmali 54 vzoriek vína zo štyroch rôznych oblastí Španielska. Vzorka bola skúmaná spektrometrom 2 minúty, pričom bolo zaznamenaných 50 spektier (každé spektrum bolo výsledkom spriemerovania 32 skenov). Keďže nie všetky spektrá boli potrebné, autori sa rozhodli použiť iba spektrá od 15 do 49 a taktiež zachovali iba hodnoty driftového času medzi 15.4 do 24 ms. Týmto zmenšili celkovú veľkosť dátového súboru, s ktorým ďalej vedeli lepšie pracovať. V ďalšom kroku vypočítali priemer a štandardnú odchýlku pre každý driftový čas v meraní, aby vedeli vzorku vína reprezentovať jedným riadkom. Pomocou klasifikátora k-NS boli schopný klasifikovať vzorku vína s presnosťou vyššou ako 90%.

Med je dobrým zdrojom rôznych antioxidantných, nutričných, antimikrobiálnych či protizápalových látok. Avšak so zvýšeným dopytom po mede sa zvyšuje aj množstvo podvodníkov, ktorí med pančujú. V práci od Aliaño-González, María José a kol. [1] autori pomocou iónového pohyblivostného spektrometra skúmali látky (prímesy), ktoré boli pridané do medu. Na vykonanie experimentu použili celkovo 77 vzoriek, z ktorých 2 boli čisté medy, 5 čistých prímiesí a 70 pančovaných medov. Každé meranie bolo reprezentované viacerými spektrami, avšak aby bolo meranie reprezentované iba jedným riadkom, v každom driftovom čase urobili súčet všetkých hodnôt naprieč spektrami. Vo výsledku ešte potrebovali dáta normalizovať od 0 po 1. Vďaka experimentom boli schopný potvrdiť, že iónový pohyblivostný spektrometer je rýchlou a spoľahlivou metódou na detegovanie rôznych prímiesí v mede.

V práci od Martin Sabo a kol. [3] autori zisťovali, ako by mohli malé podniky pôsobiace v oblasti kávového priemyslu efektívne a lacno vyhodnocovať kvalitu kávy. Ako potenciálne

vhodné riešenie uvádzajú meranie kvality pomocou IMS spektrometrov. Tieto zariadenia sú prenosné a cenovo dostupné. Je tu však problém interpretácie komplexných dát nameraných spektrometrom. Tá si vyžaduje profesionála z odboru. Martin Sabo a kol. preto navrhujú využitie strojového učenia na dátach zo spektrometra, čo by, v prípade úspechu, eliminovalo daný problém, pretože model by vedel vyhodnotiť automaticky kvalitu kávy. V rámci výskumu mali k dispozícii 41 vzoriek kávy. 22 vzoriek bola čistá Arabica a 8 vzoriek čistá Robusta. Zvyšných 11 vzoriek boli mixy v pomere 90% - 10% Arabica a 10% - 90% Robusta. Následne boli spektrometrom tieto vzorky odmerané a s využitím modulu strojového učenia v softvéri MaSaTECH¹ bol predikovaný pomer Arabicy a Robusty vo vytvorených mixoch. Na predikovanie bola použitá random forest architektúra. Priemerná presnosť predikcií je 96%, čo predstavuje vysoko uspokojivé výsledky. Vďaka tejto skutočnosti sa vynára tvrdenie, že prítomnosť profesionála na interpretáciu dát by nemusela byť potrebná, a teda daná metóda by mohla predstavovať jednoduché a výhodné riešenie na zisťovanie kvality kávy pre malé podniky pôsobiace v oblasti kávového priemyslu.

4.5.3. Základné informácie o dátach

Dáta používané pri riešení nášho tímového projektu sú získavané z rastlín technickej konopy. Následne sú rastliny namerané a spracované do vzoriek, ktoré prechádzajú do fázy merania iónovým spektrometrom. Parametre (teplota pri meraní, dĺžka merania,...) merania v iónovom spektrometre je kalibrované podľa našich potrieb - pre všetky merania sú rovnaké, z dôvodu minimalizovania chybovosti pri následnom porovnávaní vzoriek.

Dáta získavané zo spektrometra máme dostupné v dvoch formátoch: vo formáte .txt a vo formáte .csv. Dáta obsahujú hodnoty látok namerané v čase - hodnoty môžeme vizualizovať ako body na karteziánskej súradnicovej sústave, pričom poznáme, kde ležia aj na x-ovej, aj y-ovej osi. Meranie, ktoré môže trvať až 20 minút tvorí séria časových úsekov. Časový úsek, ktorý odzrkadľuje látky namerané za určitých podmienok, nazývame spektrum. Náš dataset teda obsahuje súbor týchto spektier patriacich pod jedno meranie.

Cieľom našej práce je analyzovať a porovnávať viaceré merania (série spektier), na základe čoho chceme vedieť čo možno najlepšie určiť, kedy je pre farmára vhodný čas pre zber rastliny z hľadiska najvyššej hodnoty vybranej látky (v našom prípade je to CBD).

4.6. Návrh riešenia dátového modelu

V návrhu riešenia nasledujeme štandardné postupy dátovej analýzy:

1. Predspracovanie dát
2. Trénovanie modelu
3. Vyhodnotenie modelu

V rámci predspracovania dát plánujeme vykonať nasledovné kroky:

¹ MaSaTECH control software; <https://www.masatech.eu/control-software>

- skontrolovanie chýbajúcich hodnôt (v našom prípade to môžu byť aj chýbajúce spektrá v jednotlivých meraniach)
- detekcia a nahradenie outlierov
- výber relevantných úsekov spektier
- výpočet črt (možná agregácia spektier do jedného riadku)
- normalizácia
- výber črt

Model budeme trénovať na rôznych reprezentáciách dát:

- dáta vo forme vektorov
- agregácia spektier vypočítaním črt: priemer, medián, štandardná odchýlka,...

Pri trénovaní nevynecháme hyperparametrizáciu formou GridSearch.

Na konci vyhodnotíme úspešnosť modelov pomocou metrík:

- mean squared error (MSE)
- root mean squared error (RMSE)
- mean absolute error (MAE)

4.7. Implementácia dátového modelu

Ku dňu odovzdania sa žiaľ stále nepodarilo získať potrebné vzorky, a preto nemáme skutočné dáta, s ktorými by sme vedeli pracovať. Avšak, aby sme nezostali stáť na mieste a začali postupne vypracovávať projekt, od vedúceho sme dostali vzorku konopného čaju, nad ktorou sme vedeli naimplementovať niektoré časti.

4.7.1. Spracovanie .txt súborov a vytvorenie .csv

Skôr, ako budeme vedieť pracovať s dátami, potrebujeme údaje dostať z textových súborov do vhodnej podoby. V textových súboroch sú jednotlivé spektrá uložené pod sebou ako dvojica hodnôt oddelených medzerou, pričom každé spektrum obsahuje hlavičku, ktorá definuje poradové číslo spektra, teplotu, tlak a iné veličiny popisujúce aktuálne spektrum a podmienky, v ktorých bolo zaznamenané.

Textový súbor sme po načítaní spracovali vo funkciách tak, že sme ním prechádzali po jednom riadku a ukladali sme si hodnoty do poľa, z ktorého sme neskôr vytvorili dataframe a ten sme uložili ako súbor s príponou csv. Po konzultácií s vedúcim sme sa taktiež dohodli, že je pre nás bezpredmetné ukladať si driftový čas (prvý stĺpec z textového súboru) a teda sme si ukladali iba druhý stĺpec zo súboru. Tieto hodnoty sme si v našom projekte ukladali do nových stĺpcov s názvom y_X , kde X je číslo od 1 po 668 (každé spektrum bolo reprezentované 668-timi hodnotami). Výsledná transformácia textového súboru na csv súbor je zobrazená na obrázku nižšie.

```

Spectrum 1.
Drift field intensity: 579.837 [V/cm]
Pressure: 797.616 [mbar]
Temperature T1: 100.622 [C]
Drift tube length: 10.82 [cm]
Analysis time: 0.15 [sec]
mobility
21.47508      0.063
20.64911      0.071
19.88433      0.062
19.17418      0.062
18.51300      0.060
17.89590      0.059
17.31861      0.058
16.77741      0.061

```



measurement_id	drift_field_intensity	pressure	temperature	drift_tube_len	analysis_time	y_1	y_2	y_3	y_4
1.0	579.837	797.616	100.622	10.82	0.15	0.063	0.071	0.062	0.062
1.0	579.837	797.616	100.622	10.82	0.30	0.065	0.072	0.061	0.063
1.0	579.837	797.616	100.690	10.82	0.45	0.063	0.071	0.061	0.062
1.0	579.837	797.616	100.759	10.82	0.60	0.063	0.071	0.059	0.062
1.0	579.837	797.616	100.690	10.82	0.75	0.063	0.071	0.060	0.062

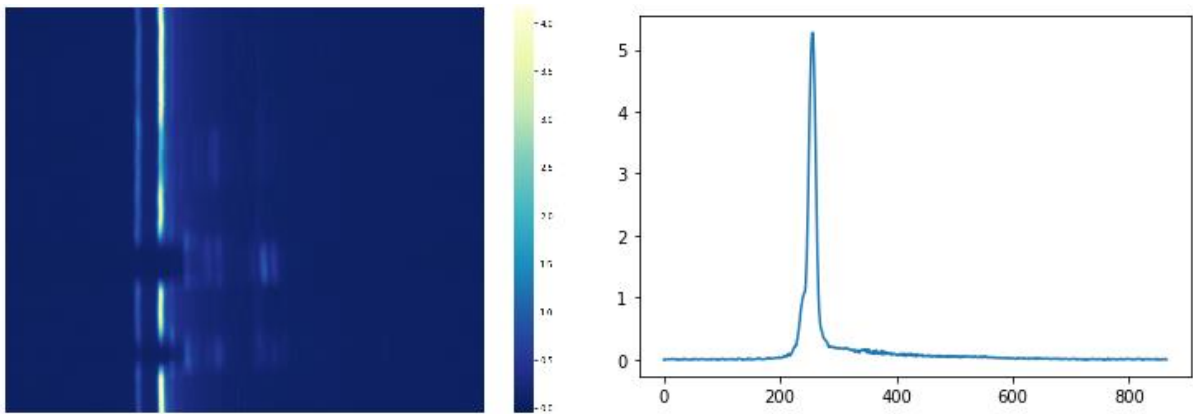
4.7.2. Unifikácia dát

Je potrebné skontrolovať, či boli počas merania zaznamenané všetky spektrá a ak nie, je treba do DataFrame-u so spektrami doplniť chýbajúce spektrá. Táto kontrola je realizovaná na základe hodnôt v stĺpci `analysis_time`. `analysis_time` prvého spektra určuje zároveň `analysis_time` offset hodnotu medzi jednotlivými spektrami. Ak je medzi `analysis_time` dvoch susedných spektier v DataFrame-e väčší rozdiel ako offset hodnota, znamená to, že chýba jedno alebo viac spektier. Chýbajúce spektrá sú doplnené nasledujúcim spôsobom: všetky hodnoty v riadku okrem `measurement_id` a `days_to_harvest` (tie sú v jednom meraní rovnaké pre všetky spektrá) sú nastavené na NaN. NaN hodnoty sú po pridaní chýbajúcich spektier doplnené pomocou funkcie `pandas.DataFrame.interpolate`².

4.7.3. Prvotná vizualizácia dátového súboru

V ďalšom kroku sme sa pokúsili dáta vizualizovať na rôznych typoch grafov, aby sme vedeli povedať ako sú rozložené a taktiež nám vizualizácie slúžili na kontrolu, že dáta sú správne a teda neobsahujú neplatné hodnoty. Použili sme čiarový graf a teplotnú mapu. Ukážky sú na obrázku nižšie.

² <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.interpolate.html>



4.7.4. Odstránenie šumu v dátach

V rámci časti odstránenia šumu v dátach sme sa venovali metódam predspracovania dát zo spektrometra s dôrazom na odstránenie šumu. V prvej časti sme sa zaoberali prehľadovým štúdiom týchto metód, v druhej ich implementáciou.

Študovali sme metódy:

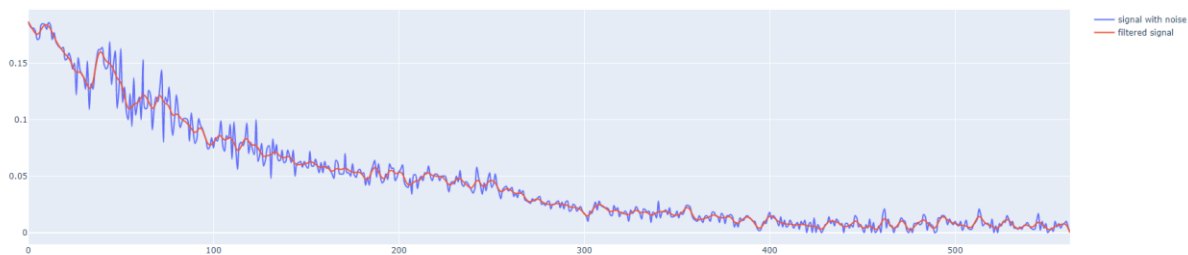
- Low pass a High pass filter
- Fourier Transform
- Wavelet Transform
- Kalman filter
- Recursive Least Square (RLS)
- Least Mean Square Error (LMS)
- Wiener-Kolmogorov Filters

Na základe prehľadového štúdia metód a základnej implementácie na vzorke dát sme identifikovali metódy Low pass filter a High pass filter ako vhodné na použitie v našom prípade predspracovania dát. Ostatné zo skupiny študovaných metód sme uzavreli ako nevhodné pre riešenie nášho projektu resp. možné ďalšieho štúdia resp. Rozhodli sme sa postupovať smerom použitia metód s čo najmenším modifikovaním dát.

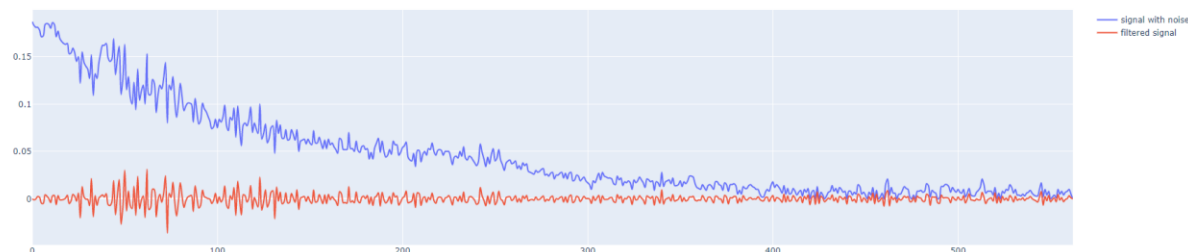
Implementovali sme metódy:

- Low pass filter
- High pass filter

Vo fáze odstránenia šumu v dátach sme sa zaoberali aj otázkou "odstránenia reakčného peak-u", ktorý v našich dátach predstavuje prvé zachytené látky iónovým spektrometrom. Nakoľko môže byť potrebný na určenie a spracovanie ďalších zachytených látok sme sa rozhodli dáta nemodifikovať jeho odstránením.



Použitie: Low pass filter



Použitie: High pass filter

4.7.5. Agregácie meraní do jedného riadku

Jednou z reprezentácií dát, ktorú chceme v projekte použiť, je agregovanie spektier konkrétneho merania do jedného riadku. Pre tieto účely sme použili funkcie, ktoré nám poskytuje knižnica pandas. Sú nimi funkcie *min()*, *max()*, *mean()*, *median()*, *std()*, *var()*, *mad()*, *sum()* a taktiež funkcie pre vypočítanie asymetrickosti a špicatosti (*skewness* a *kurtosis*). Agregáčnne funkcie boli vždy použité pre konkrétny stĺpec, teda celkovo sme vypočítali agregáčnne funkcie pre 668 stĺpcov. Vytvorili sme tak nový dataframe s počtom stĺpcov 6680 (668*10).

4.7.6. Porovnávanie vektorov

Podopierajúc návrh ďalšieho postupu sme zvolili prístupy porovnania vektorov na overenie možného smerovania projektu postaveného na báze porovnávaní meraní. Venovali sme sa nasledujúcim dvom skupinám prístupov:

- Similarity Based Metrics:
 - Pearson's correlation
 - Spearman's correlation
 - Kendall's Tau
 - Cosine similarity
 - Jaccard similarity - iba na binárne použitie (nepredstavuje náš prípad)
- Distance Based Metrics:
 - Euclidean distance
 - Manhattan distance

Prístupy "Similarity Based Metrics" pri použití na porovnanie vektorov resp. spektier ukázali menšie rozdiely medzi jednotlivými spektrami. Konkrétne boli výsledky medzi porovnaním jedného spektra s ostatnými z merania v intervale od 1 - ~0.9 pre metódy "Pearson's correlation", "Spearman's correlation" a "Cosine similarity", v intervale od 1 - ~0.8 pre metódu "Kendall's Tau". Výsledky z prístupov "Distance Based Metrics" boli v intervale 0 - ~0.25 pre metódu "Euclidean distance" a 0 - ~4.2 pre metódu "Manhattan distance". Metóda "Manhattan distance" sa tak ukázala ako metóda s najväčším rozsahom rozdielnosti, následkom čoho sa jej chceme ďalej venovať.

4.7.7. Výber črt

Hoci ešte nemáme dáta, ale pripravili sme si funkcie na výber najlepších črt spomedzi všetkých. Skúšali sme najmä filtrovacie metódy na základe korelačnej štatistiky a vzájomnej informácie. Použili sme knižnicu *sklearn*, z ktorej sme použili funkciu *SelectKBest()*, pomocou ktorej nám program bol schopný nájsť k najlepších črt. Ako parameter do funkcie sme vložili *f_regression* pre korelačnú štatistiku a *mutual_info_regression* pre vzájomnú informáciu. K sme volili postupne od 100 po polovicu všetkých (v našom prípade 3300), pričom sme k zväčšovali v každej iterácii o 100 v prípade *f_regression* a o 200 v prípade *mutual_info_regression*. Výsledky sme validovali pomocou modelu lineárnej regresie, pričom ako evaluačnú metriku sme použili MSE (mean square error). Výsledky MSE pre každé k sme zobrazili na grafe, aby sme videli ako klesá chyba so stúpajúcim k .

Pokúšali sme sa pripraviť funkcie na výber črt wrapper a vnorenými metódami, ale na dátovom súbore s dummy dátami neboli funkcie možné overiť.

5. Referencie

- [1] Aliaño-González, María José, et al. *Novel method based on ion mobility spectroscopy for the quantification of adulterants in honeys*. 114, Food Control, 2020.
- [2] Garrido-Delgado, R., et al. *Direct coupling of a gas–liquid separator to an ion mobility spectrometer for the classification of different white wines using chemometrics tools*. 471-479, Talanta, 2011.
- [3] Kateryna Trach, Oleksandr Prystopiuk, Martin Sabo, and Štefan Matejčík. *Preliminary results on quantitative gc-ims analysis of arabica and robusta coffees in mixes*. 2019.