

Dokumentace k závaznému výsledku projektu

Projekt:

Inteligentní robotická ochrana zdraví ekosystému hydroponického skleníku

FW01010381

Výstup:

Software pro predikci sklizně

T A
Č R

Tento projekt je spolufinancován se státní podporou Technologické agentury ČR a Ministerstva průmyslu a obchodu v rámci **Programu TREND**.

www.tacr.cz

www.mpo.cz

Obsah

UŽIVATELSKÁ PŘÍRUČKA	3
SPUŠTĚNÍ A PREDIKCE	3
WEBOVÝ FORMULÁŘ	3
PARAMETRY FORMULÁŘE	4
TECHNICKÁ DOKUMENTACE	6
POŽADAVKY:	6
IMPLEMENTACE FUNKČNÍCH ČÁSTÍ	6
<i>Databázový model</i>	6
<i>Predikční model</i>	6
<i>Webová aplikace</i>	6
<i>Možnosti rozšíření</i>	6
POUŽITÉ ALGORITMY A TECHNOLOGIE	7
POPIS VÝSLEDKŮ	7

Uživatelská příručka

Demonstrační webová aplikace a predikční model jsou vytvořeny v programovacím jazyce python s využitím knihoven *Streamlit* (webová aplikace), *scikit-learn*, *pickle*, *pandas* a *numpy* (predikční model). Pro spuštění webové aplikace je tedy potřeba, aby na dané platformě byl funkční python interpreter včetně knihoven *Streamlit* a *pickle*. Samotná webová aplikace potom umožňuje získat predikci výnosu rajčat v kg/m² po zadání údajů o sadbě (odrůda, počet dnů od zahájení sadby, výnos z posledního týdne) a interních a externích enviromentálních datech (osvětlení, srážky, teplota, sluneční osvit).

Spuštění a predikce

V případě splnění požadavků na spuštění (python interpreter, *Streamlit* a *pickle*) je možné webovou aplikaci spustit následujícím příkazem (z adresáře obsahujícího skript `web_app.py` a modely GB1, GB2, DecisionTree1, DecisionTree2, RF1 a RF2):

```
streamlit run web_app.py
```

Po spuštění je webová aplikace dostupná na adrese:

http://IP_ADDRESS:8501

kde IP_ADDRESS reprezentuje jednoznačnou IP adresu serveru/počítače na kterém je SW spuštěn.

Webový formulář

Pro získání predikce výnosu na následující dva týdny je potřeba vyplnit webový formulář:

The screenshot displays the Streamlit web application interface. On the left, under the 'Input' section, there are several fields: 'Variety' (a dropdown menu set to 'Sweetelle'), 'Week_num' (a text input set to '11'), 'Lights' (a slider set to 1), 'Days_since_start' (a text input set to '221'), 'Rain_mean' (a text input set to '4.50'), 'Rain_STD' (a text input set to '15.20'), 'Rain_min' (a text input set to '0'), 'Rain_max' (a text input set to '108.00'), and 'Temp_mean' (a text input set to '6.02'). On the right, there are five more text input fields: 'Irradiance_mean' (174.04), 'Irradiance_STD' (286.98), 'Irradiance_min' (0.00), 'Irradiance_max' (1092.00), and 'Yield_m2' (0.56). Below these inputs, the 'Prediction' section shows results for three models: 'Gradient Boosted Trees', 'Decision Trees', and 'Random Forest'. Each model displays two predictions: 'Next week' and 'The week after next week', with values in kg/m² and a real value in parentheses. For example, Gradient Boosted Trees predicts 0.6805218361122623 kg/m² for the next week and 0.5658675898387793 kg/m² for the week after next, with a real value of 0.671749.

Změnou hodnot ve formuláři automaticky dojde k přepočtu predikce pro nově zadané hodnoty.

Pro správnou funkčnost formuláře je nutné vyplnit všechny položky!

Parametry formuláře

Požadované parametry vycházejí z dostupných dat získaných v rámci projektu a jejich definice je následující:

Parametr	Popis	Typ	Hodnoty
Variety	Odrůda	Výčtová hodnota	Marinice, Sweetelle, Duelle, Other
Week_num	Kalendářní týden	Celočíselná hodnota	1-52
Lights	Umělé osvětlení	Celočíselná hodnota	1/0 (Ano, Ne)
Days_since_start	Počet dní od začátku sadby	Celočíselná hodnota	0-x
Rain_mean	Průměrné srážky (mm/h)	Kladná reálná hodnota	0-x
Rain_STD	Směrodatná odchylka srážek (mm/h)	Kladná reálná hodnota	0-x
Rain_min	Minimální srážky (mm/h)	Kladná reálná hodnota	0-x
Rain_max	Maximální srážky (mm/h)	Kladná reálná hodnota	0-x
Temp_mean	Průměrná teplota (°C)	Reálná hodnota	
Temp_STD	Směrodatná odchylka teploty (°C)	Reálná hodnota	
Temp_min	Minimální teplota (°C)	Reálná hodnota	
Temp_max	Maximální teplota (°C)	Reálná hodnota	
Irradiance_mean	Průměrný osvit (W/m2)	Kladná reálná hodnota	0-x
Irradiance_STD	Směrodatná odchylka osvitu (W/m2)	Kladná reálná hodnota	0-x
Irradiance_min	Minimální osvit (W/m2)	Kladná reálná hodnota	0-x
Irradiance_max	Maximální osvit (W/m2)	Kladná reálná hodnota	0-x
Yield_m2	Sklizeň v aktuálním týdnu (kg/m2)	Kladná reálná hodnota	0-x

*Všechny hodnoty jsou založeny na dostupných datech projektu. Vztažná jednotka pro hodnoty je 1 týden, tedy průměry, směrodatné odchylky, minima a maxima jsou uváděna v tomto rozlišení. U enviromentálních ukazatelů se jedná o údaje poskytované každou celou hodinu.

Technická dokumentace

Požadavky:

- Napojení na různé zdroje dat – enviromentální systém skleníku ISII + LetsGrow, předpovědní model ÚCEEB, interní záznamy o sklizni
- Propojení datových zdrojů do databázové aplikace
- Predikce sklizně na následující 2 týdny
- Rozhraní pro využití agronomy při predikci

Implementace funkčních částí

Databázový model

Pro účely shromažďování dat byla vytvořena MySQL databáze, do které jsou pomocí PHP skriptů pravidelně integrována a ukládána data z různých datových zdrojů, ať už s využitím dostupných API (LetsGrow, ÚCEEB) nebo vlastního řešení. Databázová struktura byla navržena pro usnadnění vyhledávání a zpracování dat získaných z externích zdrojů.

Predikční model

Predikční model byl vytvořen v jazyce Python s využitím knihoven *scikit-learn*, *numpy*, *pandas* a *statistics*. Vzniklý predikční model je serializován pomocí modulu *pickle* tak, aby bylo možné jej nasadit i mimo vývojovou platformu, např. v produkčním prostředí.

Webová aplikace

Demonstrační webová aplikace byla vytvořena pomocí Python knihovny *Streamlit*, která umožňuje rychlou tvorbu především formulářových aplikací, takže je velmi vhodná i pro demonstraci využití predikčního modelu a je uživatelsky přívětivá např. pro agronomy, kteří ji mohou využít pro pomoc při predikci.

Možnosti rozšíření

Vzniklý systém je modulární a tak je možné jej dále rozšiřovat pomocí úpravy konkrétních modulů. Pokud by byl např. dostupný nový datový zdroj, je potřeba tento datový zdroj propojit s databázovým modelem a integrovat nově dostupná data. Stejně tak je možné upravit predikční model pro inkorporaci nových dat nebo použít model zcela nový za předpokladu zachování napojení do databázového modelu. Webová aplikace pak reflektuje požadovaná vstupní data predikčního modelu a poskytuje zobrazení jeho výstupu, takže je libovolně rozšiřitelná/upravitelná za předpokladu zachování těchto požadavků.

Použité algoritmy a technologie

V rámci řešení tohoto projektového cíle byly využity následující technologie a algoritmy:

- Python
 - Scikit-learn
 - Pandas
 - Pickle
 - Statistics
 - Numpy
 - Streamlit
- PHP
- MySQL
- Algoritmy
 - Gradient Boosted Trees
 - Decision Trees
 - Random Forest

Popis výsledků

V rámci řešení tohoto dílčího cíle projektu byla identifikována klíčová data, která je potřeba vhodně sbírat, ukládat a zpracovávat. Tato data lze rozdělit do tří kategorií:

- Data o produkci
- Data interních charakteristik
- Data externích vlivů

Dále bylo věnováno úsilí designu datového modelu a také specifikaci jako modulárnosti s ohledem na unikátní metodiky pro rozličné skleníky/farmy, a tedy nutnosti generalizace unikátnosti prediktorů formou abstrakce. Byla také nastolena otázka synchronizace ručního zadávání a automatizovaných vstupů dat. Dále byly ve spolupráci s farmou Bezdínek identifikována data s nejvýznamnějším vlivem na predikci produkce, a to:

- měření rajčete (spolu s informací o počtu kusů na střepci),
- fáze rostliny,
- typ odrůdy,
- typ prostředí (nasvěcovaný/nenasvěcovaný),
- množství slunečního záření (externě).

Byly definovány také oblasti modulárního systému, díky kterým bude možné dosáhnout zpřesnění predikce pomocí následujících modulů a vstupních dat:

- obrazová data z robota (a následné analýzy těchto obrazových dat),
- detekce škůdců,
- data o počasí/klimatu.

Bylo nutné taky zvažovat omezení ve formě sběru v různé fázi rostliny a dále budoucí otázky typů poskytnutých vizualizací dat pro farmu.

Pro účely ukládání dat a možnost jejich následného zpracování byla v prvotní fázi řešení projektu na datové infrastruktuře NWT a.s. vytvořena OLAP kostka, do které se postupně

napojovaly jednotlivé datové zdroje ze všech tří výše zmiňovaných hlavních kategorií. Pro zjednodušení strojového zpracování dat produkce byl stávající nestrukturovaný systém ukládání těchto údajů nahrazen strukturovaným sdíleným dokumentem, který je vhodný k následnému automatizovanému zpracování, ale zároveň uživatelsky přívětivý. Data z tohoto dokumentu byla automaticky nahrávána do OLAP kostky. V rámci řešení bylo také prováděno prověřování různých engineů vyhledávání a datové analýzy (např. Elasticsearch), ze kterých by bylo možné veškerá data vyčítat pro účely zpracování predikčními algoritmy a jejich výsledky validovány.

Na farmě Bezdínek je pro sledování interních charakteristik skleníku využíván systém LetsGrow, který poskytuje potřebné informace pro predikci stresových faktorů a jejich dopad na produkci rostlin. V prvotní fázi řešení projektu bylo zvažováno napojení tohoto systému do výše zmíněné OLAP kostky. Nakonec vzniklo komplexnější řešení - viz. níže. Velké úsilí projektového týmu bylo tedy právě vloženo do získání přístupu k datům v systému LetsGrow. Tato data však bohužel nejsou jednoduše přístupná, a tak proběhla série jednání s producentem data-analytické platformy LetsGrow.com B.V., holandskou společností Hoogendoorn. Výsledkem jednání zástupců všech tří zapojených subjektů v tomto projektu (NWT, UTB a farma Bezdínek) byla dohoda o zpřístupnění dat ze systému LetsGrow, které by měly být původně dostupné v Q1 2022, nakonec byly k dispozici až v průběhu Q3/2022. Toto mělo za následek opožděnější vývoj predikční platformy. Dále byly specifikovány důležité interní faktory a jejich vliv na produkci rostlin, které se v LetsGrow systému nezaznamenávají a bylo potřeba s nimi pracovat v dalších fázích řešení tohoto dílčího cíle.

Jedním z nejdůležitějších faktorů při predikci sklizně jsou také externí meteorologické charakteristiky, především informace o intenzitě slunečního záření. Farma Bezdínek je vybavena vlastní meteostanicí poskytující základní sadu dat, která je zaznamenávána ve výše zmíněném systému LetsGrow. Při predikci sklizně jsou ovšem běžně pracovníky farem využívány i externí datové zdroje s predikcí vývoje počasí v následujících dnech a tato data nebyla ve stávajícím stavu nikde ukládána. Pro zautomatizování procesu bylo tedy potřeba zajistit zdroj predikčních dat. Toto se podařilo díky navázání spolupráce s Univerzitním Centrem Energeticky Efektivních Budov ČVUT v Praze (UCEEB), které se dlouhodobě zabývá predikcí slunečního osvětlení v oblasti solární energetiky. Jejich historická data osvětlení spolu s predikčními daty pro oblast farmy Bezdínek byla prvotně napojena do OLAP kostky a připravena tak pro další zpracování. Jako další možný vhodný zdroj dat byl zkoumán archiv předpovědních dat Evropského centra pro střednědobé předpovědi počasí (ECMWF – 10 dní), který byl v průběhu roku 2020 veřejně zpřístupněn, včetně specifických SW nástrojů pro čtení datových sad, anebo Global Forecast systém (GFS - 16 dní), zpřístupněný společností NOAA. Data z těchto dvou globálních předpovědních modelů nakonec nebyla použita.

Dále v pozdějších fázích řešení tohoto dílčího cíle byla prohlubována spolupráce s týmy zpracovávajícími obrazová data ze skleníku (detekce rajčat) a pro detekci škůdců, tedy za účelem zvažovaného propojování jednotlivých modulů datového modelu.

Významným milníkem při řešení tohoto dílčího cíle bylo zajištění napojení a integrity co největšího množství dostupných datových zdrojů. V datové infrastruktuře NWT a.s. byly nakonec propojena veškerá historická data o produkci ze skleníku, data z LetsGrow systému, historická data osvětlení spolu s predikčními daty pro oblast farmy Bezdínek přístupná přes API od UCEEB, a dále nová data o produkci získávaná z farmy Bezdínek pomocí nových interaktivních formulářů. Bylo tak zajištěna tvorba komplexního datasetu pro predikční a validační úlohu. Přičemž v záloze stále zůstává původně vytvořená OLAP datová kostka.

Výstupem je prototyp predikčního modelu v jazyce Python, navržený pro zpracování výše uvedeného souboru komplexních dat. Především z důvodů opožděného zajištění přístupu dat z LetsGrow a postupné redukci variace pěstovaných odrůd na farmě Bezdínek, dosahuje prototyp modelu aktuálně nižší úrovně generalizace (co se týče komplexní predikce např. ve skleníku pěstující řadu různých odrůd produktu(ů)). Přesnost a generalizaci modelu je samozřejmě možné zvýšit při dostatku dat za delší časové období, jejich variaci, případně napojení dalších datových zdrojů např. u potenciálního uživatele – společnosti vlastníci několik skleníků.

Základní statistické údaje nejúspěšnějších predikčních modelů pracujících s limitovaným datovým zdrojem jsou uvedeny následující tabulce.

Parametr	Predikce 1 týden [%]	Predikce 2 týdny [%]
Minimum	60.73	59.89
25%	86.06	90.49
Medián	101.67	99.10
75%	121.96	114.40
Maximum	202.67	187.34
Průměr	105.61	103.56
Směrodatná odchylka	25.92	24.30

Pro účely ukázky využití modelu pro predikci na základě aktuálních dat byla vytvořena webová aplikace pomocí knihovny *Streamlit* v Pythonu a její výstup je prezentován na následujícím obrázku.

Input

Variety: Sweetelle

Week_num: 11

Lights: 0 to 1

Days_since_start: 221

Rain_mean: 4.50

Rain_STD: 15.20

Rain_min: 0

Rain_max: 108.00

Temp_mean: 6.02

Prediction

Gradient Boosted Trees

Next week: 9.5865218361122829 kg/m2 (real: 0.671749)

The week after next week: 9.5658675698383793 kg/m2 (real: 0.666545)

Decision Trees

Next week: 9.6717493891882271 kg/m2 (real: 0.671749)

The week after next week: 9.564134948128951 kg/m2 (real: 0.666545)

Random Forest

Next week: 9.7328252245881359 kg/m2 (real: 0.671749)

The week after next week: 9.5628866739689265 kg/m2 (real: 0.666545)