

Projekt č. FW01010381

Inteligentní robotická ochrana zdraví ekosystému hydroponického skleníku

Popis výsledku projektu dokazující dosažení výsledku - Systém pro predikci a analýzu stavu rostlin v rámci skleníku

Tento dokument popisuje naplnění cíle projektu a postup dosažení výsledku projektu. Data, jednotlivé kódy a funkční algoritmy jsou uloženy u hlavního řešitele projektu.

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 Bezdíněk identifikována data s nejvýznamnějším vlivem na predikci produkce, a to:

- měření rajčete (spolu s informací o počtu kusů na střípci),
- 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 zpracovávající 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 potencionálního uživatele – společnosti vlastnící několik skleníků.

Základní statistické údaje nejúspěšnějších predikčních modelů pracujících s limitovaným datovým zdrojem jsou uvedeny v tabulce (Tab. 1). 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 obrázku (Obr. 1).

Tab.1: Statistické údaje predikce sklizně (na 1 a 2 týdny dopředu) s využitím nejúspěšnějších modelů (Gradient Boosted Trees) pro leave-one-out cross-validation.

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

Input

Variety
Sweetelle

Week_num
11

Lights
0

Days_since_start
221

Rain_mean
4.50

Rain_STD
15.20

Rain_min
0

Rain_max
108.00

Temp_mean
6.02

Irradiance_mean
174.04

Irradiance_STD
286.98

Irradiance_min
0.00

Irradiance_max
1092.00

Yield_m2
0.56

Prediction

Gradient Boosted Trees

Next week: 0.6865218361122829 kg/m2 (real: 0.671749)

The week after next week: 0.5658675698383793 kg/m2 (real: 0.666545)

Decision Trees

Next week: 0.6717493891882271 kg/m2 (real: 0.671749)

The week after next week: 0.564134948128951 kg/m2 (real: 0.666545)

Random Forest

Next week: 0.7328252245881359 kg/m2 (real: 0.671749)

The week after next week: 0.5628866739688265 kg/m2 (real: 0.666545)

Obr. 1 Demonstrační webová aplikace - využití predikčních modelů pro nová data.