Akademia Łomżyńska

Wydział nauk Informatyczno-Technologicznych



**Dokumentacja projektowa**

Opracowanie przygotowane na zajęcia z przedmiotu

**Wydziałowy projekt zespołowy**

**Zespół autorski**

Gracjan Eryk Penk

Damian Piotrowski

**Prowadzący zajęcia projektowe**

dr inż. Janusz Rafałko

Informatyka

Studia stacjonarne I stopnia, rok IV, semestr VII

Rok akademicki: 2024/2025

Spis treści

[Wstęp 3](#_Toc180092680)

[1. Wstępne założenia 4](#_Toc180092681)

[1.1. Podział zadań 4](#_Toc180092682)

[1.2. Wybór języka programowania 4](#_Toc180092683)

[1.3. Dobór narzędzi do przetwarzania danych 4](#_Toc180092684)

[1.4. Przegląd literatury 5](#_Toc180092685)

[1.4.1. FOOD DEMAND PREDICTION USING MACHINE LEARNING 5](#_Toc180092686)

[1.4.2. A case analysis of a sustainable food supply chain distribution system—A multi-objective approach 6](#_Toc180092687)

[Bibliografia 6](#_Toc180092688)

Wstęp

Głównym celem projektu jest stworzenie modelu prognozowania zapotrzebowania na produkty spożywcze, wykorzystując dane historyczne oraz techniki uczenia maszynowego. Prognozy te mogą być stosowane do optymalizacji zarządzania zapasami, redukcji kosztów logistycznych i ograniczenia marnotrawstwa żywności. Projekt ma na celu opracowanie narzędzia, które pozwoli na przewidywanie ilości zamawianych produktów na podstawie danych sprzedażowych i innych czynników wpływających na popyt, takich jak sezonowość, promocje czy lokalne wydarzenia.

Prognozowanie zapotrzebowania na żywność jest istotnym problemem dla wielu firm z sektora spożywczego, ponieważ wpływa na zarządzanie łańcuchem dostaw, koszty operacyjne oraz minimalizację strat związanych z nadprodukcją lub niedoborami produktów. Celem projektu jest zbudowanie modelu predykcyjnego, który będzie analizował dane historyczne, wykrywał wzorce i prognozował przyszły popyt. Model ten zostanie opracowany z wykorzystaniem narzędzi takich jak regresja, sieci neuronowe, LSTM (Long Short-Term Memory) czy modele szeregów czasowych (np. ARIMA, Prophet).

# Wstępne założenia

## Podział zadań

Osoba 1:

* Przygotowanie danych do pracy.
* Wstępna implementacja modeli predykcyjnych oraz algorytmów uczenia maszynowego
* Współpraca przy tworzeniu ostatecznego modelu.
* Przygotowanie dokumentacji.

Osoba 2:

* Stworzenie ostatecznego modelu.
* Przeprowadzenie walidacji i optymalizacji modeli.
* Analiza wyników i ich interpretacja.
* Prezentacja wyników pracy.

## Wybór języka programowania

Projekt zostanie zrealizowany w języku Python, który jest szeroko stosowany w analizie danych oraz uczeniu maszynowym. Python oferuje bogaty ekosystem narzędzi, które ułatwią analizę, przetwarzanie danych oraz budowę modeli predykcyjnych.

## Dobór narzędzi do przetwarzania danych

Wstępna propozycja doboru narzędzi do przetwarzania danych:

* Pandas – do analizy i manipulacji danymi.
* NumPy – do operacji numerycznych.
* Scikit-learn – do implementacji modeli klasycznych, jak regresja, drzewa decyzyjne.
* TensorFlow/Keras – do trenowania sieci neuronowych i głębokiego uczenia (w tym LSTM do predykcji szeregów czasowych).
* Prophet – do zaawansowanego modelowania szeregów czasowych, z uwzględnieniem sezonowości i trendów.
* Matplotlib/Seaborn – do wizualizacji danych i wyników modelowania.

## Przegląd literatury

### FOOD DEMAND PREDICTION USING MACHINE LEARNING

Artykuł „Food Demand Prediction Using Machine Learning” opublikowany w International Research Journal of Engineering and Technology (IRJET) opisuje metody prognozowania popytu na produkty spożywcze w restauracjach za pomocą algorytmów uczenia maszynowego i analizy statystycznej. Celem pracy jest przewidywanie liczby zamówień, co pomaga restauracjom zarządzać zapasami składników o krótkim terminie przydatności, zmniejszając tym samym marnotrawstwo żywności i poprawiając efektywność operacyjną.

Artykuł wskazuje, że prognozowanie popytu na żywność jest trudnym zadaniem z powodu licznych czynników wpływających na zmienność zamówień, takich jak sezonowość, preferencje klientów, promocje czy pogoda. Celem artykułu jest opracowanie modeli, które pozwolą przewidywać liczbę zamówień w restauracjach, co pomoże w efektywniejszym zarządzaniu zapasami.

W badaniu wykorzystano różne techniki uczenia maszynowego, w tym:

* Bayesian Linear Regression: Model oparty na sieciach Bayesowskich, który używa rozkładów prawdopodobieństwa do prognozowania zmiennych. Model ten umożliwia przewidywanie na podstawie zależności warunkowych między zmiennymi losowymi.
* Random Forest: Technika wykorzystująca wiele drzew decyzyjnych do prognozowania. Model ten dobrze radzi sobie zarówno z danymi liniowymi, jak i nieliniowymi.
* Support Vector Machine (SVM): Popularny algorytm do klasyfikacji, który tworzy hiperpłaszczyznę w celu rozdzielenia danych na różne klasy.
* LASSO Regression: Technika regresji stosująca „kurczenie” zmiennych, co pozwala na uproszczenie modelu i wybór najistotniejszych zmiennych.
* XGBoost: Zaawansowany algorytm oparty na gradient boosting, który jest wyjątkowo wydajny pod względem obliczeń i często daje lepsze wyniki w porównaniu do innych technik.

W badaniu przeprowadzono prognozowanie liczby zamówień na podstawie danych wewnętrznych (takich jak liczba zamówień z poprzednich tygodni) oraz danych zewnętrznych (np. promocje). Stwierdzono, że algorytm XGBoost osiągnął najlepsze wyniki, zapewniając najwyższą dokładność prognozowania. Inne algorytmy, takie jak regresja Bayesowska i LASSO, miały niższą dokładność.

Dodatkowo, artykuł podkreśla znaczenie inżynierii cech, czyli procesu przekształcania danych w formy, które są bardziej użyteczne dla algorytmów uczenia maszynowego. Na przykład, użyto kodowania etykiet dla danych kategorycznych.

Autorzy zauważają, że dokładność modeli można jeszcze poprawić, uwzględniając dodatkowe czynniki, takie jak zwyczaje kulturowe czy święta religijne. W przyszłości planowane jest rozszerzenie metod na inne branże, takie jak prognozowanie zapotrzebowania na pracowników.

Artykuł pokazuje, że zastosowanie algorytmów uczenia maszynowego w prognozowaniu popytu na żywność może znacznie poprawić dokładność przewidywań. Zastosowanie algorytmu XGBoost okazało się najskuteczniejsze, ale istnieje potencjał na dalsze ulepszenia poprzez uwzględnienie dodatkowych danych kontekstowych.

### A CASE ANALYSIS OF A SUSTAINABLE FOOD SUPPLY CHAIN DISTRIBUTION SYSTEM – A MULTI-OBJECTIVE APPROACH

Badanie „A case analysis of a sustainable food supply chain distribution system—A multi-objective approach” przeprowadzone przez ‘Management Group’ ze Szkoły Biznesu Uniwersytetu Miasta Dublin dotyczy analizy zrównoważonego systemu dystrybucji w łańcuchu dostaw żywności, ze szczególnym naciskiem na przemysł mleczarski w Irlandii. Przedstawiono model dystrybucji dwuwarstwowej, którego celem jest minimalizacja emisji CO2 oraz kosztów związanych z transportem mleka.

Zielony system dystrybucji uwzględnia zarówno wpływ na środowisko, jak i efektywność ekonomiczną. Model równoważy redukcję emisji węglowych z kontrolą kosztów.

Model uwzględnia również alternatywne scenariusze, które zakładają otwarcie zamkniętych dotąd tras dystrybucji. Ma to na celu zwiększenie odporności systemu dystrybucji na różne nieprzewidziane zmiany, takie jak zamknięcie tras lub problemy operacyjne.

Wyniki modelu są także przedstawiane w formie geograficznej, co pozwala na wizualizację optymalnych tras transportowych. Geograficzne rozmieszczenie tras pomaga lepiej zrozumieć, które trasy są najbardziej zrównoważone z punktu widzenia zarówno kosztów, jak i emisji węglowych.

**Technologie wykorzystane w badaniu:**

W badaniu zastosowano trzy różne **algorytmy genetyczne** – **NSGA-II**, **MOGA-II** oraz **metodę hybrydową**. Algorytmy genetyczne to techniki inspirowane procesami ewolucji biologicznej, które są używane do znajdowania optymalnych rozwiązań w problemach z dużą liczbą zmiennych i ograniczeń, takich jak logistyka transportu. W tym przypadku, algorytmy te pomogły w identyfikacji najlepszych tras dystrybucji, które znajdują się na tzw. froncie Pareto, czyli zestawie rozwiązań, gdzie żadna z nich nie jest gorsza od innych pod względem wszystkich celów optymalizacyjnych. Wyniki analizy wykazały, że algorytm NSGA-II (ang. Non-Dominated Sorting Genetic Algorithm II) okazał się najbardziej efektywny w porównaniu z pozostałymi metodami.

**NSGA-II (Non-Dominated Sorting Genetic Algorithm II)**: Jest to jeden z najczęściej używanych algorytmów do optymalizacji wielokryterialnej. NSGA-II sortuje populację rozwiązań na podstawie ich dominacji względem innych rozwiązań – szuka tych, które nie są gorsze w żadnym z badanych aspektów (w tym przypadku, emisje i koszty). Algorytm generuje zbiór rozwiązań, który jest rozłożony na froncie Pareto, dając wiele opcji wyboru.

**MOGA-II (Multi-Objective Genetic Algorithm II)**: Jest to inna wersja algorytmu genetycznego do rozwiązywania problemów wielokryterialnych. Podobnie jak NSGA-II, MOGA-II szuka optymalnych rozwiązań na froncie Pareto, ale może różnić się w sposobie zarządzania populacją i selekcji rozwiązań.

**Metoda hybrydowa (GA + Sequential Quadratic Programming)**: Jest to kombinacja algorytmów genetycznych z bardziej tradycyjnymi metodami optymalizacji numerycznej, takimi jak programowanie kwadratowe sekwencyjne (SQP). Podejście hybrydowe często stosuje się, aby połączyć zalety obu metod – eksploracyjnych właściwości GA z precyzją optymalizacji numerycznej.

**TOPSIS (Technique for Order Preference by Similarity to Ideal Solution)**: Jest to narzędzie wielokryterialne służące do porównywania różnych opcji decyzyjnych. Umożliwia ocenę, które rozwiązanie jest najbliższe do "idealnego" (najlepszego możliwego), bazując na odległości od hipotetycznego idealnego punktu w wielowymiarowej przestrzeni decyzji. W badaniu TOPSIS pomógł ocenić trasy pod kątem emisji CO2 i kosztów, tworząc hierarchię najbardziej efektywnych tras.

To badanie podkreśla rosnącą potrzebę wprowadzania bardziej ekologicznych i wydajnych systemów dystrybucji w łańcuchach dostaw żywności. Zastosowane algorytmy genetyczne i technologie optymalizacyjne stanowią skuteczne narzędzia, które umożliwiają jednoczesne zarządzanie kosztami i emisją węglową. Rozwiązania proponowane w modelu mogą być szerzej stosowane w innych sektorach, które muszą balansować między zrównoważonym rozwojem a efektywnością operacyjną.

Bibliografia

1. International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 p-ISSN: 2395-0072 Volume: 07 Issue: 06, June 2020, Pages 3672-3675
2. International Journal of Production Economics Volume 152,June 2014,Pages 71-87
3. https://www.sciencedirect.com/science/article/abs/pii/S0925527314000437