Obraz zawierający Czcionka, tekst, Grafika, logo

Opis wygenerowany automatycznie

**Sztuczna inteligencja w biznesie**

***NARZĘDZIE ESTYMACJI WYSOKOŚCI KREDYTU NA PODSTAWIE DANYCH FINANSOWYCH KLIENTÓW***

|  |  |
| --- | --- |
| **Prowadzący:**  dr inż. Rafał Gasz | **Autorzy:**  *Paweł Cedzich*  *Paweł Siemiginowski*  *Jan Kłek Marek Pietrowicz* |
| *Informatyka*  *Specjalność: Systemy inteligentne* | |
| Rok akademicki 2024/2025 | |

**SPIS TREŚCI**

[1. IDENTYFIKACJA PROBLEMU BIZNESOWEGO 3](#_Toc188635789)

[1.1. ANALIZA POTRZEB 3](#_Toc188635790)

[1.2. WYZNACZENIE CELÓW 4](#_Toc188635791)

[2. ZBIERANIE I ANALIZA DANYCH 5](#_Toc188635792)

[2.1. ZBIERANIE DANYCH 5](#_Toc188635793)

[2.2. CZYSZCZENIE DANYCH 5](#_Toc188635794)

[2.3. EKSPLORACYJNA ANALIZA DANYCH (EDA) 7](#_Toc188635795)

[3. WYBÓR PODEJŚCI AI 8](#_Toc188635796)

[3.1. OKREŚLENIE METODOLOGII 8](#_Toc188635797)

[3.2. WYBÓR ALGORYTMÓW 8](#_Toc188635798)

[3.3. MODELOWANIE REGUŁ BIZNESOWYCH 10](#_Toc188635799)

[4. TWORZENIE PROTOTYPU (PoC) 11](#_Toc188635800)

[4.1. BUDOWA WSTĘPNEGO MODELU 11](#_Toc188635801)

[4.2. TESTOWANIE PoC 11](#_Toc188635802)

[5. TRENOWANIE I WALIDACJA MODELU 13](#_Toc188635803)

[5.1. TRENOWANIE MODELU 13](#_Toc188635804)

[5.2. WALIDACJA 13](#_Toc188635805)

[5.3. POPRAWKI I OPTYMALIZACJA 14](#_Toc188635806)

[6. IMPLEMENTACJA I INTEGRACJA Z SYSTEMAMI BIZNESOWYMI 16](#_Toc188635807)

[6.1. INTEGRACJA MODELU 16](#_Toc188635808)

[6.2. SKALOWANIE 17](#_Toc188635809)

[7. WNIOSKI 19](#_Toc188635810)

1. **IDENTYFIKACJA PROBLEMU BIZNESOWEGO**

W dzisiejszym dynamicznie rozwijającym się świecie finansów, instytucje bankowe  
i kredytowe stają przed wyzwaniem efektywnego zarządzania ryzykiem kredytowym. Kluczowym problemem biznesowym, który narzędzie estymacji wysokości kredytu ma rozwiązać, jest precyzyjne określenie zdolności kredytowej klientów na podstawie ich danych finansowych. Wprowadzenie takiego narzędzia ma na celu automatyzację procesu oceny kredytowej, co pozwoli na szybsze i bardziej dokładne podejmowanie decyzji kredytowych.

* 1. **ANALIZA POTRZEB**

Analiza potrzeb wskazuje, że organizacje finansowe dążą do optymalizacji operacji poprzez redukcję czasu i kosztów związanych z manualną oceną zdolności kredytowej. Automatyzacja tego procesu nie tylko przyspieszy obsługę klientów, ale również zminimalizuje ryzyko błędów ludzkich, co w efekcie poprawi doświadczenia klientów. Dodatkowo, precyzyjne narzędzie estymacyjne pozwoli na lepsze zarządzanie portfelem kredytowym, co jest kluczowe dla stabilności finansowej instytucji. W kontekście rosnącej konkurencji na rynku finansowym, szybka i dokładna ocena zdolności kredytowej może stanowić istotną przewagę konkurencyjną.

Wprowadzenie narzędzia estymacji wysokości kredytu na podstawie danych finansowych klientów ma również na celu poprawę transparentności procesu kredytowego. Klienci będą mieli możliwość lepszego zrozumienia, na jakiej podstawie podejmowane są decyzje kredytowe, co może zwiększyć ich zaufanie do instytucji finansowej. Ponadto, narzędzie to może przyczynić się do lepszego zarządzania ryzykiem kredytowym poprzez identyfikację potencjalnych zagrożeń na wcześniejszym etapie.

W kontekście regulacji prawnych i wymogów dotyczących ochrony danych, narzędzie to musi być zaprojektowane w sposób zapewniający pełną zgodność z obowiązującymi przepisami. Obejmuje to zarówno ochronę danych osobowych klientów, jak i zapewnienie, że procesy decyzyjne są przejrzyste i sprawiedliwe. Wdrożenie takiego narzędzia wymaga również odpowiedniego szkolenia personelu, aby zapewnić jego efektywne wykorzystanie.

* 1. **WYZNACZENIE CELÓW**

Wyznaczenie celów projektu jest niezbędne do oceny jego sukcesu. Kluczowe wskaźniki sukcesu (KPI) obejmują dokładność prognozowania maksymalnej kwoty kredytu oraz minimalnej liczby miesięcy wymaganych do jego spłaty. Dodatkowo, istotnym wskaźnikiem będzie czas potrzebny na przetworzenie danych i wygenerowanie rekomendacji kredytowej. Ostatecznym celem jest stworzenie narzędzia, które będzie nie tylko dokładne, ale również elastyczne, umożliwiając dostosowanie do specyficznych potrzeb różnych instytucji finansowych. Warto również podkreślić, że narzędzie to ma potencjał do integracji  
z istniejącymi systemami bankowymi, co dodatkowo zwiększy jego wartość użytkową.

Podsumowując, narzędzie estymacji wysokości kredytu na podstawie danych finansowych klientów ma na celu automatyzację i optymalizację procesu oceny zdolności kredytowej, co przyniesie korzyści zarówno instytucjom finansowym, jak i ich klientom. Kluczowe wskaźniki sukcesu obejmują dokładność prognoz, czas przetwarzania danych oraz elastyczność narzędzia. Wprowadzenie tego rozwiązania może znacząco poprawić efektywność operacyjną  
i konkurencyjność instytucji finansowych na rynku.

1. **ZBIERANIE I ANALIZA DANYCH**
   1. **ZBIERANIE DANYCH**

W celu stworzenia narzędzia estymacji wysokości kredytu, niezbędne było zgromadzenie odpowiednich danych finansowych klientów. Dane te zostały pobrane z publicznie dostępnego zbioru danych na platformie Kaggle, zawierającego informacje dotyczące ryzyka finansowego związanego z zatwierdzaniem kredytów. Zbiór danych obejmował różnorodne zmienne, takie jak wiek, roczny dochód, status zatrudnienia, poziom wykształcenia, doświadczenie zawodowe, stan cywilny, liczba osób na utrzymaniu, miesięczne zobowiązania finansowe, historia bankructw, saldo kont oszczędnościowych i bieżących, całkowite aktywa i pasywa, miesięczny dochód, staż pracy, wartość netto, miesięczna rata kredytu, kwota kredytu oraz czas trwania kredytu.

* 1. **CZYSZCZENIE DANYCH**

Proces czyszczenia danych był kluczowy dla zapewnienia dokładności i spójności zbioru danych. W pierwszym etapie usunięto duplikaty oraz uzupełniono brakujące wartości. Następnie przeprowadzono analizę zmiennych, aby zidentyfikować te, które mają największy wpływ na zdolność kredytową klientów. W wyniku tej analizy zdecydowano się na usunięcie niektórych kolumn, które nie wnosiły istotnych informacji lub były redundantne. Ostatecznie, do dalszej analizy wybrano następujące zmienne:

* Age - wiek klienta
* AnnualIncome - roczny dochód
* EmploymentStatus - status zatrudnienia
* EducationLevel - poziom wykształcenia
* Experience - doświadczenie zawodowe (w latach)
* MaritalStatus - stan cywilny
* NumberOfDependents - liczba osób na utrzymaniu
* MonthlyDebtPayments - miesięczne zobowiązania finansowe
* BankruptcyHistory - historia bankructw
* SavingsAccountBalance - saldo konta oszczędnościowego
* CheckingAccountBalance - saldo konta bieżącego
* TotalAssets - całkowite aktywa
* TotalLiabilities - całkowite pasywa
* MonthlyIncome - miesięczny dochód
* JobTenure - staż pracy (w latach)
* NetWorth - wartość netto
* MonthlyLoanPayment - miesięczna rata kredytu

Na końcu wyznaczone zostały dwie zmienne, które będą stanowić cel do wyznaczenia przez przyjęty algorytm:

* LoanAmount - kwota kredytu (*zmienna docelowa*)
* LoanDuration - czas trwania kredytu (*zmienna docelowa*)

Wszystkie wybrane zmienne zostały sprawdzone pod kątem brakujących wartości. Proces ten polegał na dokładnym przeanalizowaniu każdej kolumny w celu upewnienia się, że nie zawiera ona brakujących danych. W przypadku wykrycia brakujących wartości, przeprowadzono odpowiednie działania, takie jak uzupełnianie braków lub usuwanie niekompletnych rekordów.

Dodatkowo, dane tekstowe musiały zostać zmapowane na wartości cyfrowe, aby mogły być użyte w modelach AI. Przykładowy kod mapowania wartości tekstowych na liczbowe przedstawiono na rysunku 2.2.1.

Obraz zawierający tekst, zrzut ekranu, Czcionka, linia

Opis wygenerowany automatycznie  
Rys. 2.2.1. Mapowanie danych tekstowych na wartości liczbowe.

* 1. **EKSPLORACYJNA ANALIZA DANYCH (EDA)**

Eksploracyjna analiza danych (EDA) została przeprowadzona w celu zrozumienia struktury danych, zmiennych oraz potencjalnych zależności między nimi. EDA pozwoliła na identyfikację trendów i anomalii, które mogą być istotne dla modelu. Na przykład, analiza korelacji między zmiennymi wykazała, że roczny dochód oraz wartość netto mają silny wpływ na maksymalną kwotę kredytu, jaką klient może otrzymać. Z kolei zmienne takie jak liczba osób na utrzymaniu oraz miesięczne zobowiązania finansowe mają istotny wpływ na minimalną liczbę miesięcy wymaganych do spłaty kredytu.

W trakcie analizy danych zwrócono również uwagę na rozkład wartości poszczególnych zmiennych oraz ich wpływ na zmienne docelowe, czyli LoanAmount i LoanDuration.  
W ramach EDA przeprowadzono również szczegółowe sprawdzenie wartości zerowych oraz typów danych, aby upewnić się, że wszystkie zmienne są odpowiednio przygotowane do modelowania [rys. 2.3.1].

Obraz zawierający tekst, zrzut ekranu, numer

Opis wygenerowany automatycznie  
Rys. 2.3.1. Wynik działania funkcji sprawdzania wartości zerowych.

1. **WYBÓR PODEJŚCI AI**
   1. **OKREŚLENIE METODOLOGII**

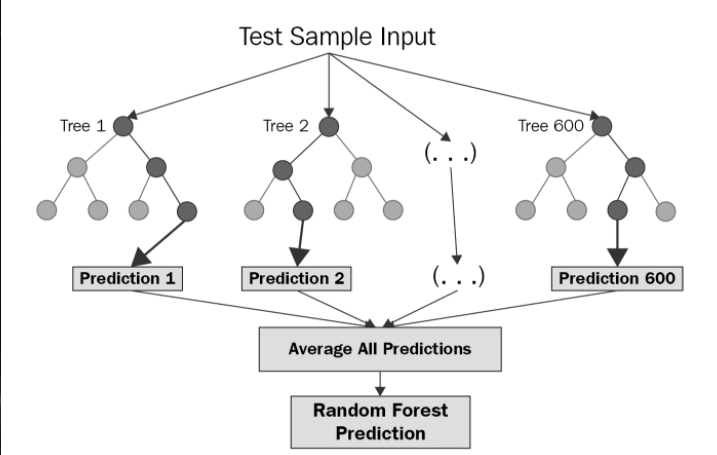
Wybór odpowiedniej technologii AI jest niezwykle istotny dla skutecznego rozwiązania problemu estymacji wysokości kredytu na podstawie danych finansowych klientów. W tym projekcie zdecydowano się na zastosowanie metod uczenia maszynowego (machine learning), które są szczególnie efektywne w analizie dużych zbiorów danych i prognozowaniu wartości liczbowych. Uczenie maszynowe pozwala na tworzenie modeli, które mogą uczyć się na podstawie danych historycznych i przewidywać przyszłe wartości z wysoką dokładnością. Dzięki temu możliwe jest uzyskanie precyzyjnych prognoz, które są niezbędne do podejmowania trafnych decyzji kredytowych.

Uczenie maszynowe oferuje szeroki wachlarz technik i algorytmów, które mogą być dostosowane do specyficznych potrzeb projektu. Wybór odpowiedniej metodologii wymagał analizy dostępnych opcji i oceny ich przydatności w kontekście estymacji wysokości kredytu. W tym przypadku, zdecydowano się na podejście oparte na regresji, które jest dobrze dopasowane do problemów związanych z prognozowaniem wartości liczbowych. Regresja pozwala na modelowanie zależności między zmiennymi wejściowymi a zmiennymi docelowymi, co jest kluczowe dla dokładnego przewidywania kwoty kredytu i okresu spłaty.

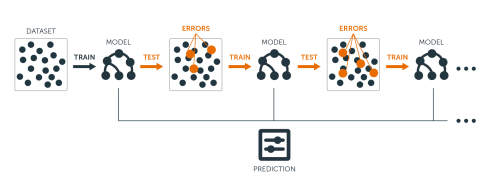
* 1. **WYBÓR ALGORYTMÓW**

W ramach metodologii uczenia maszynowego wybrano konkretne algorytmy, które najlepiej odpowiadają specyfice problemu. Do modelowania wykorzystano *MultiOutputRegressor* z biblioteki *sklearn.multioutput*, który umożliwia jednoczesne prognozowanie dwóch zmiennych docelowych: maksymalnej kwoty kredytu (LoanAmount) oraz minimalnej liczby miesięcy wymaganych do jego spłaty (LoanDuration). *MultiOutputRegressor* pozwala na efektywne zarządzanie problemami wielowymiarowymi, co jest niezwykle ważne w kontekście tego projektu.

Do trenowania modeli wybrano dwa algorytmy z biblioteki *sklearn.ensemble*: *RandomForestRegressor* oraz *GradientBoostingRegressor*. *RandomForestRegressor* jest algorytmem opartym na drzewach decyzyjnych, który tworzy wiele drzew decyzyjnych i łączy ich wyniki, aby uzyskać bardziej dokładne prognozy. Jest to algorytm odporny na przeuczenie i dobrze radzi sobie z danymi o dużej zmienności. Dzięki swojej strukturze, *RandomForestRegressor* [rys. 3.2.1] jest w stanie uchwycić złożone zależności między zmiennymi, co przekłada się na wysoką jakość prognoz.

  
Rys. 3.2.1. RandomForestRegressor – schemat działania predykcji.  
[Źródło: https://levelup.gitconnected.com/random-forest-regression-209c0f354c84]

*GradientBoostingRegressor* [rys. 3.2.2], z kolei, jest algorytmem, który iteracyjnie buduje model, minimalizując błąd prognozy. Jest szczególnie efektywny w przypadku danych złożonych i pozwala na uzyskanie wysokiej dokładności prognoz. *GradientBoostingRegressor* działa poprzez tworzenie sekwencji modeli, z których każdy kolejny model koryguje błędy poprzedniego. Dzięki temu, algorytm ten jest w stanie stopniowo poprawiać jakość prognoz, co jest niezwykle przydatne w kontekście estymacji wysokości kredytu.

  
Rys. 3.2.2. GradientBoostingRegressor – schemat działania predykcji.

[Źródło: https://vagifaliyev.medium.com/a-hands-on-explanation-of-gradient-boosting-regression-4cfe7cfdf9e]

* 1. **MODELOWANIE REGUŁ BIZNESOWYCH**

W procesie modelowania uwzględniono również specyficzne reguły biznesowe, które model AI musi spełniać. Przykładowo, polityki kredytowe banków mogą wymagać, aby maksymalna kwota kredytu była ograniczona do określonego procentu wartości netto klienta. Dodatkowo, minimalna liczba miesięcy spłaty kredytu może być uzależniona od miesięcznych zobowiązań finansowych klienta oraz jego rocznego dochodu. Uwzględnienie tych reguł biznesowych jest niezbędne dla zapewnienia, że model AI będzie generował prognozy zgodne z rzeczywistymi wymaganiami instytucji finansowych.

Narzędzie to ma służyć zarówno klientom, jak i pracownikom instytucji finansowych. Dla klientów oznacza to oszczędność czasu i możliwość samodzielnego uzyskania wstępnej oceny zdolności kredytowej. Dzięki temu, klienci mogą szybko i łatwo dowiedzieć się, na jaką kwotę kredytu mogą liczyć oraz jaki będzie minimalny okres jego spłaty. Dla pracowników natomiast, narzędzie to stanowi wsparcie w postaci wstępnych prognoz, na podstawie których mogą podejmować bardziej świadome decyzje kredytowe. Wstępne prognozy generowane przez model AI mogą być wykorzystywane jako punkt wyjścia do dalszej analizy i oceny ryzyka kredytowego, co pozwala na bardziej efektywne zarządzanie portfelem kredytowym.

1. **TWORZENIE PROTOTYPU (PoC)**
   1. **BUDOWA WSTĘPNEGO MODELU**

Na etapie tworzenia prototypu modelu AI, celem było sprawdzenie, jak dobrze rozwiązanie działa na mniejszej skali. W tym celu zbudowano wstępne modele przy użyciu dwóch algorytmów: *RandomForestRegressor* oraz *GradientBoostingRegressor*. Prototyp miał na celu przetestowanie wstępnych hipotez dotyczących zdolności kredytowej klientów oraz oceny, który z algorytmów lepiej sprawdza się w kontekście prognozowania maksymalnej kwoty kredytu (LoanAmount) oraz minimalnej liczby miesięcy wymaganych do jego spłaty (LoanDuration).

* 1. **TESTOWANIE PoC**

Przeprowadzono wstępne testy na małej próbce danych, obejmującej 1000 rekordów, aby zidentyfikować potencjalne problemy i ocenić skuteczność rozwiązania. Testy miały na celu porównanie wyników uzyskanych przez oba algorytmy pod kątem dokładności prognoz oraz szybkości działania.

Wyniki testów dla *RandomForestRegressor* były następujące:

* MSE Loan Amount: 18 528 324,12
* MSE Loan Duration: 1 660,10
* MAE Loan Amount: 15 594,77
* MAE Loan Duration: 57,14

Wyniki testów dla *GradientBoostingRegressor* były następujące:

* MSE Loan Amount: 17 898 967,84
* MSE Loan Duration: 1 596,29
* MAE Loan Amount: 15 369,37
* MAE Loan Duration: 55,78

*Mean Squared Error* (MSE) i *Mean* *Absolute* *Error* (MAE) to dwa kluczowe wskaźniki używane do oceny jakości prognoz modeli regresyjnych. MSE mierzy średnią kwadratową różnicę między prognozowanymi a rzeczywistymi wartościami, co pozwala na identyfikację dużych błędów prognozy. MAE natomiast mierzy średnią bezwzględną różnicę między prognozowanymi a rzeczywistymi wartościami, co daje bardziej intuicyjne zrozumienie przeciętnego błędu prognozy.

Wartości MSE i MAE uzyskane w testach mogą wydawać się wysokie, co wynika z kilku czynników. Po pierwsze, zbiór danych użyty do trenowania modeli nie został w pełni zweryfikowany pod względem wartości, co może wpływać na jakość prognoz. Po drugie, celem narzędzia nie jest osiągnięcie perfekcyjnej dokładności prognoz, lecz dostarczenie wstępnych ocen zdolności kredytowej, które mogą być użyteczne zarówno dla klientów, jak i pracowników instytucji finansowych. Ostateczne decyzje kredytowe i tak są podejmowane przez ludzi, którzy mogą uwzględnić dodatkowe czynniki i kontekst, którego model AI może nie uwzględniać.

Na podstawie przeprowadzonych testów zdecydowano się na wybór algorytmu *RandomForestRegressor* do praktycznej implementacji narzędzia. Pomimo że *GradientBoostingRegressor* wykazał nieco lepsze wyniki pod względem dokładności prognoz (niższe wartości MSE i MAE), *RandomForestRegressor* został wybrany ze względu na swoją większą stabilność i odporność na przeuczenie. Dodatkowo, *RandomForestRegressor* jest bardziej intuicyjny w interpretacji wyników, co ułatwia jego zastosowanie w praktyce.

Warto również zauważyć, że *RandomForestRegressor*, choć może być wolniejszy  
w działaniu w porównaniu do *GradientBoostingRegressor*, oferuje lepszą równowagę między dokładnością a interpretowalnością wyników. W kontekście narzędzia estymacji wysokości kredytu, stabilność i interpretowalność wyników są niezwykle ważne dla zapewnienia, że prognozy będą użyteczne zarówno dla klientów, jak i pracowników instytucji finansowych.

1. **TRENOWANIE I WALIDACJA MODELU**
   1. **TRENOWANIE MODELU**

Proces trenowania modelu AI rozpoczął się od podziału zebranych danych na zestaw treningowy i walidacyjny. Z 20 tysięcy rekordów, 80% (16 tysięcy rekordów) zostało przeznaczonych na trenowanie modelu, a pozostałe 20% (4 tysiące rekordów) na walidację. Trenowanie modelu polegało na dostosowaniu parametrów modelu, aby uzyskać optymalne wyniki prognozowania maksymalnej kwoty kredytu (LoanAmount) oraz minimalnej liczby miesięcy wymaganych do jego spłaty (LoanDuration).

Model został zbudowany przy użyciu *MultiOutputRegressor* z algorytmem *RandomForestRegressor*. Proces trenowania obejmował iteracyjne dostosowywanie parametrów modelu, takich jak liczba drzew w lesie losowym, aby uzyskać jak najlepsze wyniki. Model uczył się na podstawie danych historycznych, analizując zależności między zmiennymi wejściowymi a zmiennymi docelowymi.

* 1. **WALIDACJA**

Po zakończeniu trenowania modelu, jego jakość została sprawdzona na niezależnym zestawie danych walidacyjnych. Walidacja pozwala na ocenę czy model generalizuje dobrze na nowe dane, a nie tylko na te, które zostały wykorzystane do trenowania. W tym celu model został przetestowany na 4 tysiącach rekordów, które nie były używane podczas trenowania.

Wyniki walidacji dla modelu RandomForestRegressor były następujące:

* MSE Loan Amount: 58 591 705,45
* MSE Loan Duration: 524,97
* MAE Loan Amount: 4 931,50
* MAE Loan Duration: 18,07
* R2 Loan Amount: 0,6523
* R2 Loan Duration: 0,1253

*Mean Squared Error* (MSE) i *Mean Absolute Error* (MAE) to dwa kluczowe wskaźniki używane do oceny jakości prognoz modeli regresyjnych. MSE mierzy średnią kwadratową różnicę między prognozowanymi a rzeczywistymi wartościami, co pozwala na identyfikację dużych błędów prognozy. MAE natomiast mierzy średnią bezwzględną różnicę między prognozowanymi a rzeczywistymi wartościami, co daje bardziej intuicyjne zrozumienie przeciętnego błędu prognozy.

R2 (współczynnik determinacji) jest miarą, która wskazuje, jak dobrze model wyjaśnia zmienność danych. Wartość R2 bliska 1 oznacza, że model dobrze dopasowuje się do danych, natomiast wartość bliska 0 wskazuje na słabe dopasowanie. W przypadku tego modelu, wartości R2 dla Loan Amount i Loan Duration są stosunkowo niskie, co sugeruje, że model nie wyjaśnia w pełni zmienności danych. Może to wynikać z kilku czynników, takich jak niesprawdzony pod względem wartości zbiór danych oraz fakt, że celem narzędzia nie jest osiągnięcie perfekcyjnej dokładności prognoz, lecz dostarczenie wstępnych ocen zdolności kredytowej. Ostateczne decyzje kredytowe i tak są podejmowane przez ludzi, którzy mogą uwzględnić dodatkowe czynniki i kontekst, którego model AI może nie uwzględniać.

* 1. **POPRAWKI I OPTYMALIZACJA**

Na podstawie wyników walidacji przeprowadzono dalsze dostosowanie modelu, aby zoptymalizować jego działanie. Proces ten obejmował dostrajanie hiperparametrów, takich jak liczba drzew w lesie losowym (*n\_estimators*). Po przeprowadzeniu serii eksperymentów, ostatecznie ustalono wartość *n\_estimators* na 100, co zapewniło najlepszą równowagę między dokładnością prognoz a stabilnością modelu. Dzięki iteracyjnemu podejściu do optymalizacji, model został dostosowany tak, aby jak najlepiej spełniał wymagania biznesowe i dostarczał użytecznych prognoz zdolności kredytowej. Na rysunku 5.3.1. przedstawiono efekt prognozowania modelu.

Obraz zawierający linia, Wykres, Czcionka, tekst

Opis wygenerowany automatycznie

Obraz zawierający linia, Wykres, Czcionka, tekst

Opis wygenerowany automatycznieRys. 5.3.1. Prognoza modelu RandomForestRegressor przy n\_estimators = 100.

1. **IMPLEMENTACJA I INTEGRACJA Z SYSTEMAMI BIZNESOWYMI**
   1. **INTEGRACJA MODELU**

Wdrożenie rozwiązania AI w istniejące systemy IT i procesy biznesowe jest kluczowym etapem, który pozwala na praktyczne wykorzystanie modelu w codziennej działalności. W tym projekcie zastosowano IPython do prostego zintegrowania modelu w formularz kliencki. IPython umożliwia tworzenie interaktywnych formularzy, które są łatwe w obsłudze zarówno dla klientów, jak i pracowników instytucji finansowych.

Dla każdej komórki z danymi stworzono pole umożliwiające wpisanie wartości. Poprzednio zmapowane wartości tekstowe zostały przedstawione w formie list wybieralnych, co ułatwia użytkownikom wprowadzanie danych. Natomiast kwoty i długości kredytów, które klient chce uzyskać od banku, są w formie suwaków. Ich minimum i maksimum dostosowuje się dynamicznie do wpisanych wartości, co jest przedstawione na rysunku 6.1.1.

Obraz zawierający tekst, zrzut ekranu, Czcionka, numer

Opis wygenerowany automatycznieObraz zawierający tekst, zrzut ekranu, Czcionka, numer

Opis wygenerowany automatycznie  
Rys. 6.1.1. Formularz narzędzia predykcyjnego opartego na modelu RandomForestRegressor.

Po uruchomieniu walidacji formularza następuje podanie maksymalnej zdolności kredytowej klienta, z uwzględnieniem oprocentowania, które w tym przypadku wynosi 5%. Formularz wyświetla również minimalną ilość miesięcy spłaty dla maksymalnej kwoty kredytu oraz maksymalną ratę miesięczną, która jest obliczana jako kwota kredytu podzielona przez liczbę miesięcy. Poniżej tych wartości wyświetla się kwota, która została wybrana przez klienta oraz ilość miesięcy spłaty. Pod tymi dwoma wartościami jest automatycznie wyliczona kwota raty, którą klient będzie musiał płacić. To jest przedstawione na rysunku 6.1.2.

Obraz zawierający tekst, Czcionka, zrzut ekranu

Opis wygenerowany automatycznie  
Rys. 6.1.2. Podsumowanie działania formularza; na górze zdolność kredytowa,  
na dole wyliczenia do wybranego przez klienta kredytu.

Proces walidacji formularza weryfikuje, czy klient nie przekracza swojej zdolności kredytowej. W przypadku wystąpienia takiego błędu pojawia się informacja o konieczności modyfikacji danych bądź kontaktu z konsultantem. Kod odpowiedzialny za walidację formularza jest przedstawiony na rysunku 6.1.3.

Obraz zawierający tekst, zrzut ekranu, Czcionka

Opis wygenerowany automatycznie  
Rys. 6.1.3. Fragment kodu walidacji wartości formularza.

* 1. **SKALOWANIE**

Rozwiązanie AI może być zaimplementowane w różne miejsca w organizacji, w zależności od jej infrastruktury i potrzeb. Może to obejmować aplikacje webowe, aplikacje mobilne lub systemy stacjonarne. Skalowanie rozwiązania na pełną organizację pozwala na obsługę dużej ilości danych i procesów, co jest niezbędne w przypadku instytucji finansowych.

Warto podkreślić, że narzędzie wymaga wcześniejszego przetrenowania na danych bankowych danej organizacji. Użycie bardziej realnych danych specyficznych dla konkretnej instytucji finansowej powinno jeszcze bardziej polepszyć działanie predykcji i dostarczyć bardziej precyzyjnych prognoz zdolności kredytowej. Dzięki temu narzędzie będzie mogło lepiej spełniać wymagania biznesowe i dostarczać wartościowych informacji zarówno dla klientów, jak i pracowników.

1. **WNIOSKI**

Projekt narzędzia estymacji wysokości kredytu na podstawie danych finansowych klientów wykazał, że zastosowanie metod uczenia maszynowego może znacząco usprawnić proces oceny zdolności kredytowej. Wykorzystanie algorytmu *RandomForestRegressor* w połączeniu z *MultiOutputRegressor* pozwoliło na stworzenie modelu, który jest w stanie prognozować maksymalną kwotę kredytu oraz minimalną liczbę miesięcy wymaganych do jego spłaty.

Proces zbierania i analizy danych, a także ich czyszczenie i eksploracyjna analiza danych (EDA), były kluczowe dla zapewnienia jakości i spójności zbioru danych. Wybór odpowiednich zmiennych oraz ich przetworzenie na wartości cyfrowe umożliwiły skuteczne trenowanie modelu. Wyniki walidacji modelu, choć nie idealne, wskazują na potencjał narzędzia do dostarczania użytecznych prognoz zdolności kredytowej.

Implementacja narzędzia w środowisku *IPython* umożliwiła stworzenie interaktywnego formularza klienckiego, który jest łatwy w obsłudze i dostarcza wartościowych informacji zarówno dla klientów, jak i pracowników instytucji finansowych. Proces walidacji formularza zapewnia, że prognozy są zgodne z rzeczywistymi możliwościami kredytowymi klientów,  
a w przypadku przekroczenia zdolności kredytowej, użytkownik jest informowany  
o konieczności modyfikacji danych lub kontaktu z konsultantem.

Istnieje możliwość skalowania rozwiązania na pełną organizację oraz jego integracji  
z istniejącymi systemami IT i procesami biznesowymi, co pozwala na efektywne wykorzystanie narzędzia w codziennej działalności instytucji finansowych. Wdrożenie narzędzia w różnych formach, takich jak aplikacje webowe, mobilne czy stacjonarne, umożliwia jego elastyczne dostosowanie do potrzeb konkretnej organizacji.

Zastosowanie metod uczenia maszynowego w projekcie narzędzia estymacji wysokości kredytu może znacząco poprawić proces oceny zdolności kredytowej, dostarczając wartościowych prognoz i wspierając zarówno klientów, jak i pracowników instytucji finansowych. Dalsze przetrenowanie modelu na bardziej realnych danych specyficznych dla danej organizacji powinno jeszcze bardziej polepszyć działanie predykcji i zwiększyć użyteczność narzędzia.