

**WYDZIAŁ ZARZĄDZANIA**

Katedra Informatyki Biznesowej i Inżynierii Zarządzania

Projekt dyplomowy

*Aplikacja mobilna do nauki matematyki dla szkoły podstawowej z wykorzystaniem machine learningu do rozpoznawania odpowiedzi pisanych ręcznie*

*A mobile math learning app for elementary school using machine learning to recognize handwritten answers*

Autor: *Monika Helena Etrych*

Kierunek studiów: Informatyka i Ekonometria

Opiekun pracy: *dr Beata Basiura*

Kraków, 2023 r.

Spis treści

[Wstęp 3](#_Toc123635373)

[1. Rozpoznawanie cyfr i ich interpretacja 4](#_Toc123635374)

[1.1. Interpretacja obrazów przez komputer 4](#_Toc123635375)

[1.2. Metody uczenia maszynowego do rozpoznawania cyfr 4](#_Toc123635376)

[1.3. Obróbka obrazów z cyframi 4](#_Toc123635377)

[1.4. Weryfikacja modelu do rozpoznawania cyfr 4](#_Toc123635378)

[2. Model aplikacji do nauki uczenia matematyki 4](#_Toc123635379)

[2.1. Założenia modelu aplikacji 4](#_Toc123635380)

[2.2. Wybrane technologie zastosowane do budowy aplikacji 4](#_Toc123635381)

[2.3. Wybrane metody uczenia maszynowego zastosowane w aplikacji 4](#_Toc123635382)

[2.4. Korzystanie z modelu ML poprzez API HTTP 4](#_Toc123635383)

[3. Opis implementacji i działania aplikacji 4](#_Toc123635384)

[3.1. Ekran startowy i podstawowe funkcjonalności 4](#_Toc123635385)

[3.2. Schematy, opisy poszczególnych funkcjonalności, opis kodów 4](#_Toc123635386)

[3.3. Przykłady zastosowania aplikacji w konkretnych zadaniach 4](#_Toc123635387)

[Zakończenie 5](#_Toc123635388)

[Bibliografia 6](#_Toc123635389)

[Spis tabel, wykresów i rysunków 7](#_Toc123635390)

[Aneks (kod, tabelka z danymi, treść ankiety) 8](#_Toc123635391)

# Wstęp

1-2 stron, cel pracy, ważny problem, może być hipoteza badawcza, napisać na koniec!

Problem rozpoznawania cyfr – problematyczne, bo każdy ma inny charakter, pochyłość, duże małe. Różnice kulturowe. Pisanie cyfry od góry do dołu lub od dołu.

# Rozpoznawanie cyfr i ich interpretacja

## Interpretacja obrazów przez komputer

Obrazy reprezentowane przez komputer składają z pikseli - kwadratów, które posiadają kilka wartości określających kolor. Wyróżnia się 4 podstawowe reprezentacje: binarną, monochromatyczną (ang. grayscale), RGB i RGBA.

Reprezentacja binarna oraz monochromatyczna składają się tylko z jednej wartości dla danego pixela. W przypadku reprezentacji binarnej wartość jest zerojedynkowa, gdzie 0 oznacza, że pixel ma kolor czarny, a 1 to kolor biały. W pozostałych reprezentacjach wartość jest z zakresu od 0 do 255. W reprezentacji monochromatycznej 0 oznacza kolor czarny, 255 - biały, natomiast wartości pośrednie, to odcienie szarości.

RGB zawiera trzy wartości określające kolor: czerwony, zielony i niebieski. RGBA posiada dodatkowo jedną wartość określającą poziom przezroczystości. [1]

#### Zestawy danych zawierające

MNIST, EMIST, NIST?

Co jeszcze:

CMY i CMYK?

Grafika rastowa i wektorowa?

Formaty?

TRYBY

Może obrazek obrazka w różnych stylach?

Wyróżnia się 3 podstawowe tryby koloru:

* tryb [czarno-biały](https://pl.wikipedia.org/wiki/Achromatyzm_(grafika)) – występują w nim wyłącznie białe i czarne [piksele](https://pl.wikipedia.org/wiki/Piksel),
* tryb [monochromatyczny](https://pl.wikipedia.org/wiki/Monochromatyzm_(grafika)) (tzw. skala szarości) – oprócz koloru czarnego i białego występuje cała gama jasności pośrednich (szarości),
* tryb [kolorowy](https://pl.wikipedia.org/wiki/Barwa) – piksele mają przyporządkowane kolory z określonego modelu [przestrzeni barw](https://pl.wikipedia.org/wiki/Przestrze%C5%84_barw), np. [RGB](https://pl.wikipedia.org/wiki/RGB), [CMYK](https://pl.wikipedia.org/wiki/CMYK) i inne.

## Metody uczenia maszynowego do rozpoznawania cyfr

Metody uczenia maszynowego są coraz częściej wykorzystywane do rozpoznawania cyfr w różnych dziedzinach, takich jak bankowość czy usługi pocztowe. Uczenie maszynowe pozwala na automatyczne i precyzyjne rozpoznawanie cyfr. Wiele papierowych formularzy w sposób szybki i dokładny może zostać przeniesiona do komputera. Natomiast rozpoznawanie kodów pocztowych znacząco przyspiesza proces sortowania listów i przesyłek [3].

Do rozpoznawania cyfr wykorzystuje się algorytmy uczenia nadzorowanego. W tym podejściu, dane uczące zawierają cyfry wraz z odpowiadającymi im etykietami, czyli poprawnymi wartościami cyfr. Algorytm uczony na takich danych jest w stanie nauczyć się rozpoznawać cyfry na podstawie cech wizualnych, takich jak kształt i proporcje. [1, 5]

Wśród popularnych algorytmów uczenia maszynowego stosowanych do rozpoznawania cyfr znajdują się między innymi konwolucyjne sieci neuronowe (CNN), wielowarstwowe sieci neuronowe (MLP) oraz algorytm maszyn wektorów nośnych (SVM). Zastosowanie tych algorytmów wraz z odpowiednio przygotowanymi danymi uczącymi, pozwala na uzyskanie skuteczności rozpoznawania cyfr na poziomie nawet 99% [4, 5].

Algorytm maszyn wektorów nośnych (SVM) polega na maksymalizacji odległości pomiędzy hiperpłaszczyzną, a marginesem przynależności do poszczególnych klas, poprzez zmianę położenia hiperpłaszczyzny. Istnieje wiele możliwych hiperpłaszczyzn, a jej kształt zależy od złożoności problemu - liczby klas do zaklasyfikowania. Kiedy liczba klas wynosi 2, to hiperpłaszczyzna jest prostą, a dla 3 - płaszczyzną 2D itd. [6]

Wielowarstwowy perceptron (MLP) składa się z 3 podstawowych warstw: warstwy wejściowej, ukrytych warstw oraz warstwy wyjściowej.

Wymienione metody zostały porównane pod kątem czasu wykonania oraz dokładności dla danych treningowych i testowych. Najwyższy wynik dokładności dla danych treningowych osiągnął SVM -99.98%, a dla testowych CNN - 99.31% [5].

W innym badaniu porównano duże modele wielowarstwowe sieci neuronowe (MLP) z różnymi konfiguracjami liczby warstw oraz neuronów. W rezultacie najlepszy wynik uzyskano dla MLP składającej się z ośmiu warstw i liczb neuronów: 2500, 2000, 1500, 1000, 500, 10. Proces uczenia modelu obejmował aż 2000 epok, a osiągnięta dokładność wyniosła 99,65%. Warto zauważyć, że cyfry, które zostały źle sklasyfikowane przez model, były trudne do określenia nawet dla człowieka [11].

Co jeszcze:

- jeszcze kilka odwołań do badań

- może opisać inne metody jak działają

- CNN na pewno opisać później w 2 rozdziale

## Obróbka obrazów z cyframi

W procesie rozpoznawania cyfr przez model, kluczowe znaczenie ma odpowiednie przekształcenie obrazów przed ich przekazaniem do modelu. Wstępne przetwarzanie obrazów ma na celu przygotowanie danych do dalszej analizy. Ponadto zmniejsza czas potrzebny do wytrenowania modelu i zwiększa szybkość otrzymania wyniku. Na przykład, kiedy wejściowy obraz jest stosunkowo duży, zmniejszenie rozmiaru znacząco zmniejsza czas trenowania, bez znacznej utraty dokładności modelu.

Sieci neuronowe wymagają, aby dane wejściowe miały ten sam rozmiar, ponieważ sieci te korzystają z macierzy wag, które mają ustaloną liczbę kolumn i wierszy. Przetwarzanie obrazów przez ustawienie stałych wymiarów jest jednym ze sposobów na dostosowanie rozmiaru obrazów do wymogów sieci neuronowej.

Zmiany rozmiaru obrazu można dokonać na dwa sposoby. Pierwszy z nich polega na zmniejszeniu obrazu przez wycinanie obszaru, który zawiera obraz, a drugi sposób polega na zwiększeniu rozmiaru poprzez dodanie pustych pikseli.

W sytuacji, kiedy danych treningowych jest mało, model głębokiego uczenia może ulec przeuczeniu, czyli zbytniemu dopasowaniu do danych trenujących. Przeuczenie skutkuje słabymi wynikami na nowych danych. Aby zapobiec temu problemowi, konieczna jest odpowiednia liczba danych trenujących. Można to osiągnąć stosując augmentację, czyli przekształcenia istniejących obrazów za pomocą np. losowych zmian rotacji lub jasności. Na (rys. 1.1.) w pierwszym rzędzie przedstawiono obrazy cyfr w normalnym położeniu oraz w drugim rzędzie po zastosowaniu zmian rotacji o losowy kąt z zakresu ± 30 stopni. Augmentacja jest łatwym sposobem na zwiększenie różnorodności danych trenujących i poprawę jakości modelu.

Obraz zawierający tekst

Opis wygenerowany automatycznie

Rysunek . Obrót obrazów

Te same zabiegi przetwarzania danych mogą być stosowane zarówno w ramach wstępnego przetwarzania danych, jak i augmentacji. Jednakże, wstępne przetwarzanie danych jest stosowane zarówno na danych treningowych, jak i testowych, podczas gdy augmentacja jest stosowana wyłącznie na danych treningowych po wstępnym przetworzeniu. [2]

Co jeszcze:

Do najczęstszych przekształceń należą: centrowanie, obcinanie brzegów, zamiana trybu kolorów na greyscale.

Preprocessing: Resize, Orientation

Augmentation: Grayscale, Random Flips, Random Rotations, Random Brightness and Exposure, Random Noise

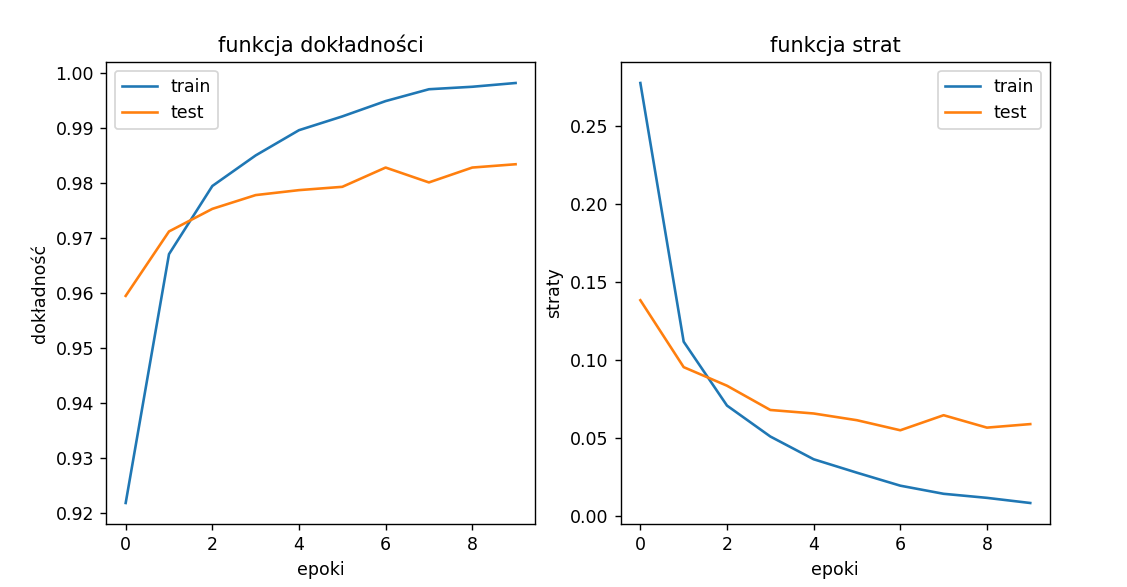
Co użyte w aplikacji

## Weryfikacja modelu do rozpoznawania cyfr

Podstawowym sposobem weryfikacji modelu uczenia maszynowego jest sprawdzenie na jakim etapie nauczenia znajduje się. Model pod wpływem czasu może być kolejno: niedouczony, wyuczony odpowiednio i przeuczony. Jednak, jeśli model jest źle skonstruowany może nie dochodzić do późniejszych etapów.

Problem niedouczenia polega na tym, że hipotezy modelu są złe i zbyt proste w porównaniu do złożoności danych. Przez to model nie opisuje wystarczająco danych. Natomiast przeuczenie polega na zbytniej złożoności modelu i w konsekwencji zbyt dokładnym opisywaniu danych treningowych i problemami w dopasowaniu nowych elementów [7].

Dwa wykresy dokładności i strat, mogą dostarczyć informacji o stanie wyuczeniu modelu. Funkcja dokładności przedstawia poziom dokładności modelu w kolejnych epokach. Tak długo jak te krzywe dla danych uczących i testujących rosną model polepsza swoją skuteczność – model nie jest jeszcze wystarczająco wyuczony. Kiedy wykres stabilizuje się, nie rośnie tak gwałtownie mamy model odpowiednio wyuczony. Jednak, gdy funkcja na zestawie danych testujących po pewnym czasie zacznie maleć – oznacza to, że model został przeuczony. Celem jest maksymalizacja współczynnika dokładności i minimalizacja strat, błędu [8].



Rysunek . Funkcja dokładności i funkcja strat

Na (rys. 1.2.) został przedstawiony model odpowiednio wyuczony. Funkcja dokładności osiąga wyższe wartości dla danych uczących niż dla danych testowych. Jest to naturalny proces, na znanych danych model osiąga lepsze wyniki niż na nowych. Funkcja strat

Interpretacja wyników, doczytać o tych różnych wskaźnikach, accuracy, i te ile dobrych dobrze klasyfikuje ile złych źle itp.

Correctly Classified

Incorrectly classified

Kappa statistic -co to? Kappa statistics value ranges from 0 to 1.  
Value 0 means totally disagreement and 1  
means full agreement.

Mean absolute error

Root Mean squared error

Relative absolute error

Root relative absolute error

dokładność, czułość, specyficzność oraz krzywe ROC i AUC.

# Model aplikacji do nauki uczenia matematyki

## Założenia modelu aplikacji

## Wybrane technologie zastosowane do budowy aplikacji

## Wybrane metody uczenia maszynowego zastosowane w aplikacji

Napisać wstęp

Konwolucyjne sieci neuronowe (ang. CNN) są często wykorzystywane do rozpoznawania obiektów na obrazach.

Konwolucyjne sieci neuronowe składają się z trzech warstw:

1. warstwa wejściowa
2. warstwa konwolucyjne
3. Filtry – które nakłada się na obrazek, wycina kawałek i oblicza się ile pixeli jest takich samych
4. Mapy cech – zbiera wyniki z obliczonych filtrów
5. warstwa aktywacji – funkcja, która przekształca mapę cech
6. warstwa łączenia (ang. pooling) – zmniejsza wymiar macierzy, AVR lub MAX
7. warstwa normalizacji wsadowej (barch normalization layer)
8. warstwa w pełni połączona – taka jak normalnie daje się do uczenia
9. warstwa wyjściowa

[9][10]

Receptive Field Size – rozmiar filtra

Stride Width – o ile pixeli przesuwamy filtr

## Korzystanie z modelu ML poprzez API HTTP

# 

# Opis implementacji i działania aplikacji

## Ekran startowy i podstawowe funkcjonalności

## Schematy, opisy poszczególnych funkcjonalności, opis kodów

## Przykłady zastosowania aplikacji w konkretnych zadaniach

# Zakończenie

Wyniki, rezultaty, udowodnienie tezy, co na przyszłość rozwinąć

# Spis tabel, wykresów i rysunków

# Wykaz rysunków

Rysunek 1.1: opracowanie własne, stworzone przy użyciu programu Python

Rysunek 1.2: opracowanie własne, stworzone przy użyciu programu Python

Rysunek 1.3: opracowanie własne, stworzone przy użyciu programu Python

# Bibliografia

[1] M. [Candocia](https://www.linkedin.com/in/max-candocia-a270b574/)*, A Simple Explanation of how Computers Recognize Images,* [Analysis for Many Audiences,](https://maxcandocia.com/)  2016, url: <https://maxcandocia.com/article/2016/Apr/06/how-computers-recognize-images/>, dostęp: 23.02.2023

[2] J. Nelson*, What is Image Preprocessing and Augmentation?,* Roboflow, 2020, url: <https://blog.roboflow.com/why-preprocess-augment/>, dostęp: 28.02.2023r.

[3] S. M. Shamim, M. B. A. Miah, A. Sarker, M. Rana, A. A. Jobair, 2018, *Handwritten Digit Recognition Using Machine Learning,* Algorithms Indonesian Journal of Science & Technology (1) vol. 3, str.: 29-39 ISSN 2528-1410, ISSN 2527-8045

# [4] M. Javed, *The Best Machine Learning Algorithm for Handwritten Digits Recognition,* 2020,

url:<https://towardsdatascience.com/the-best-machine-learning-algorithm-for-handwritten-digits-recognition-2c6089ad8f09>, dostęp: 1.03.2023

[5] R. Dixit, R. Kushwah, S. Pashine, *Handwritten Digit Recognition using Machine and  
Deep Learning Algorithms,* Artykuł Naukowy 2021, url: <https://arxiv.org/pdf/2106.12614.pdf>, dostęp: 1.03.2023

[6] R. Gandhi, *Support Vector Machine — Introduction to Machine Learning Algorithms*, 2018, url: <https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>, dostęp: 1.03.2023

[7] D. Nikolaiev *Overfitting and Underfitting Principles,* 2021,url: [*https://towardsdatascience.com/overfitting-and-underfitting-principles-ea8964d9c45c*](https://towardsdatascience.com/overfitting-and-underfitting-principles-ea8964d9c45c)*,* dostęp: *14.03.2023*

[8] J. Brownlee, *How to use Learning Curves to Diagnose Machine Learning Model Performance, 2019,* url: <https://machinelearningmastery.com/learning-curves-for-diagnosing-machine-learning-model-performance/>, dostęp: 9.03.2023

[9] J. Brownlee, *Develop Deep Learning Models on Theano and TensorFlow Using Keras,* 2016, str: 116 - 134

[10] M. Mamczur ,*Jak działają konwolucyjne sieci neuronowe (CNN)?,* 2021, url: <https://miroslawmamczur.pl/jak-dzialaja-konwolucyjne-sieci-neuronowe-cnn/>, dostęp: 14.03.2023

[11] D.C. Ciresan, U. Meier, L. M. Gambardella, J . Schmidhuber, *Deep Big Multilayer Perceptrons For Digit Recognition,* 2012, w: G. Montavon, G.B. Orr, KR. Müller, *Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science*, vol 7700, str: 581-598, Springer, Berlin, Heidelberg

GNB - <https://www.youtube.com/watch?v=H3EjCKtlVog>

Decision trees - <https://towardsdatascience.com/decision-trees-in-machine-learning-641b9c4e8052>

Random forest - <https://towardsdatascience.com/understanding-random-forest-58381e0602d2>

KNN - <https://medium.com/swlh/k-nearest-neighbor-ca2593d7a3c4>

<https://learn.g2.com/k-nearest-neighbor>

SGD - <https://towardsdatascience.com/gradient-descent-811efcc9f1d5>

# Aneks (kod, tabelka z danymi, treść ankiety)

(NOTATKI)

## Spis ludności w USA

## Rozpoznawanie japońskich znaczków xD

Podobnym zadaniem jest rozpoznawanie przez algorytm znaków specyficznych dla różnych języków np. języka chińskiego czy arabskiego.

## Akcelerometry

Data augmentation – ciekawe :D