

POLITECHNIKA WROCŁAWSKA  
WYDZIAŁ ELEKTRONIKI

---

KIERUNEK: Informatyka  
SPECJALNOŚĆ: Grafika i systemy multimedialne

**PRACA DYPLOMOWA  
MAGISTERSKA**

Model prognozowania pogody  
krótkoterminowej przy pomocy metod  
głębokiego uczenia dla wybranej lokalizacji

Deep Learning Model for Forecasting Short-  
Term Weather in Chosen Location

Autor:  
inż. Bartłomiej Edward Kramnik

Opiekun pracy:  
Dr inż. Jacek Mazurkiewicz, K30W04D03

OCENA PRACY:

---

**WROCŁAW 2020**

## Spis treści

Streszczenie .....	4
Abstract .....	4
1. Wstęp .....	5
2. Cel i zakres pracy .....	6
3. Wprowadzenie do problematyki .....	6
3.1. Elementy prognozy stanu pogodowego .....	7
3.2. Zmienność geograficzna pogody .....	8
3.3. Problem zbierania rozproszonych danych .....	9
3.4. Problem malejącej skuteczności prognoz wraz z czasem .....	9
3.5. Metody modelowania pogody .....	10
4. Przegląd istniejących rozwiązań .....	11
4.1. Global Forecast System (GFS) .....	11
4.2. Model European Centre for Medium-Range Weather Forecasts (ECMWF) .....	12
4.3. ICOSahedral Nonhydrostatic general circulation model (ICON) .....	12
4.4. Unified Model(UM) .....	12
4.5. IBM Deep Thunder .....	12
4.6. Podsumowanie .....	12
5. Wybór zbioru danych .....	13
5.1. Przegląd wybranego zbioru danych .....	13
5.2. Klimat w wybranej lokalizacji .....	15
5.3. Możliwe anomalie na zbiorze danych .....	18
5.3.1. Hipoteza: Na skutek zmian klimatu początkowe pomiary znacznie różnią się od końcowych .....	18
5.3.2. Hipoteza: Erupcja wulkanu Eyjafjallajökull na Islandii w 2010 roku, spowodowała znaczący spadek widoczności w następnych miesiącach .....	23
5.4. Podsumowanie .....	24
6. Implementacja modelu .....	24
6.1. Przegląd implementacji występujących w literaturze .....	24
6.1.1. Sieć neuronowa z jedną warstwą ukrytą (MLP) .....	24
6.1.2. Sieć rekurencyjna (RNN) .....	25
6.1.3. Sieć konwolucyjna (CNN) .....	26
6.1.4. Sieć konwolucyjna LSTM .....	26
6.1.5. Podsumowanie .....	27
6.2. Przetworzenie danych treningowych .....	27
6.2.1. Usunięcie danych niepełnych .....	29
6.2.2. Konwersja do formatu 1 na K (One hot encoding) .....	29
6.2.3. Normalizacja cech ilościowych .....	30
6.2.4. Konwersja na tensor 3 wymiarowy .....	30
6.2.5. Kompresja danych .....	31
6.2.6. Usunięcie zbędnych cech .....	32
6.3. Opracowanie struktury sieci .....	32
6.4. Funkcja straty .....	33
6.5. Uczenie sieci .....	34
6.6. Funkcje aktywacji kolejnych warstw .....	34
7. Wybór hiperparametrów modelu .....	34
7.1. Platforma testowa .....	34
7.2. Metoda testowania i ograniczenia .....	35
7.3. Wybór algorytmu optymalizacji sieci .....	35

7.4.	Wybór funkcji straty .....	37
7.5.	Wybór funkcji aktywacji .....	38
7.6.	Wybór liczby neuronów w warstwach ukrytych .....	39
7.7.	Wybór wartości początkowej współczynnika uczenia .....	41
7.8.	Podsumowanie .....	42
8.	Uczenie i testy modelu .....	42
8.1.	Uczenie modelu .....	42
8.2.	Walidacja krzyżowa .....	42
8.3.	Testy dokładności modelu w kolejnych prognozowanych godzinach .....	43
8.4.	Badanie błędu całego wektora prognozy w czasie .....	43
8.5.	Badanie prognozy opadów w czasie .....	44
8.6.	Badanie prognozy temperatury w czasie .....	45
8.7.	Badanie prognozy temperatury mokrego termometru w czasie .....	45
8.8.	Badanie prognozy temperatury w punkcie rosy .....	46
8.9.	Badanie średniego błędu prognozy wilgotności względnej .....	46
8.10.	Badanie średniego błędu prognozy ciśnienia atmosferycznego .....	47
8.11.	Badanie średniego błędu prognozy prędkości wiatru .....	47
8.12.	Badanie średniego błędu prognozy nasłonecznienia .....	48
8.13.	Badanie średniego błędu prognozy widoczności .....	48
8.14.	Badanie średniego błędu kodu synoptycznego .....	49
8.15.	Porównanie skuteczności modelu w różnych sezonach roku. ....	49
8.16.	Wnioski .....	50
9.	Podsumowanie .....	51
9.1.	Możliwe przyczyny niskiej skuteczności modelu .....	51
9.1.1.	Trudna do przewidzenia charakterystyka klimatu Irlandii .....	51
9.1.2.	Niewłaściwy dobór parametrów i architektury .....	52
9.2.	Dalsze możliwości rozwoju projektu .....	52
9.2.1.	Porównanie skuteczności modelu z innymi architekturami sieci .....	52
9.2.2.	Znalezienie lepszej metody doboru parametrów sieci .....	52
9.2.3.	Użycie modelu na innym zbiorze danych w innej lokalizacji .....	52
	Bibliografia .....	53
	Spis rysunków .....	56
	Spis tabel .....	57

## **Streszczenie**

Ludzie od zarania dziejów zajmowali się prognozowaniem pogody. Wraz z rozwojem technologii powstawały coraz skuteczniejsze metody meteorologiczne, pozwalające z większą dokładnością oszacować stan atmosfery w przyszłości. Obecnie, dzięki dynamicznemu wzrostowi mocy obliczeniowej komputerów coraz popularniejsze stają się metody wykorzystujące uczenie maszynowe i głębokie uczenie.

W ramach niniejszej pracy opracowano model głębokiego uczenia pozwalający na przeprowadzanie prognoz krótkoterminowych na obszarze Irlandii. w pierwszym etapie dokonano dogłębnej analizy problemu i rozważono istniejące rozwiązania. Następnie zdefiniowano wymagania co do zbioru danych i wybrano najlepiej im odpowiadający. W dalszych etapach dokonano wyboru architektury i parametrów modelu. Na końcu dokonano szczegółowych testów i przedstawiono wnioski wyciągnięte na w czasie realizacji projektu.

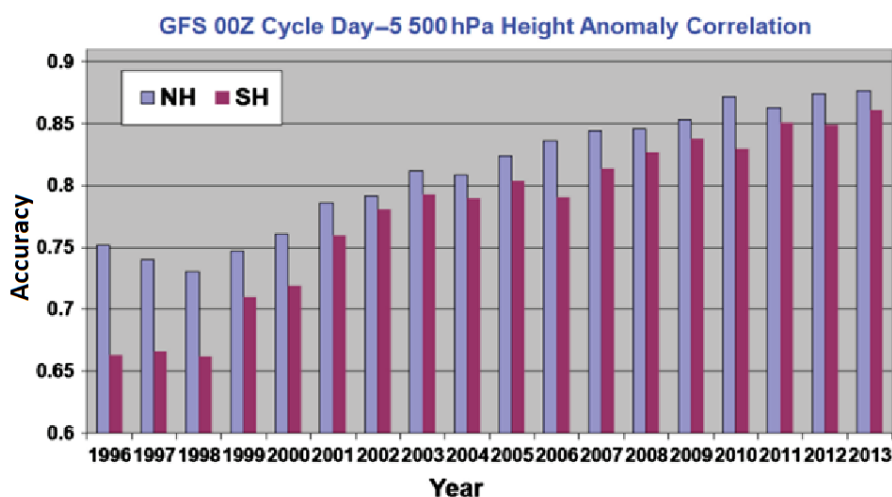
## **Abstract**

People have been dealing with weather forecasting since the dawn of time. Along with the development of technology, more and more effective meteorological methods were created, allowing more accurate estimation of the state of the atmosphere in the future. Currently, in consequence to the dynamic development of computing power, methods using machine learning and deep learning are becoming more and more popular.

This thesis contains development of deep learning model for providing short term weather forecast in the area of Ireland. At the beginning an in-depth analysis of the problem was made and existing solutions considered. Next, the data set requirements were defined and the one best suited to them was selected. In further stages, the architecture and model parameters were selected. Finally, profound tests and conclusions drawn during the project implementation are presented.

## 1. Wstęp

Prognozowaniem stanu pogody ludzkość zajmuje się od tysięcy lat, jednak jest to o tyle skomplikowany problem, że do dzisiaj nie jesteśmy w stanie skutecznie przewidzieć stanu pogody na okres dłuższy niż kilka dni w przód [1]. Chociaż współczesne modele wciąż pozostawiają wiele do życzenia, faktem jest że obecne metody stają się coraz lepsze i dokładniejsze. Historyczne modele synoptyczne, obecnie nie są już praktycznie użytkowane, a na ich miejsce pojawiają się wymagające dużej mocy obliczeniowej modele numeryczne. Coraz więcej ludzi korzysta także z dedykowanych usług jak „AccuWeather” czy „Pogoda Google” – które przetwarzają ogromną ilość danych synoptycznych z całego świata, aby wygenerować szczegółową prognozę dla każdego miejsca na ziemi.



Rys. 1.1 Zmiana skuteczności prognozy pogody 5 dniowej w kolejnych latach na przykładzie systemu GFS, gdzie: NH – półkula północna, SH – półkula południowa [2]

Jak można zauważyć na rysunku 1.1, skuteczność prognoz wyraźnie się poprawia, osiągając nawet 90% od roku 2010. Widoczna różnica pomiędzy półkulami, wynika prawdopodobnie ze specyfiki systemu – w innych rozwiązaniach nie występuje.

Moc komputerów cały czas rośnie, więc zdawałoby się, że powinniśmy osiągnąć już dawno zadowalające wyniki. Tak się jednak nie dzieje i w najbliższej przyszłości prawdopodobnie nie stanie. Stan pogody jest wciąż zbyt skomplikowany dla modeli matematycznych [1], mimo że obecnie ponad 15% najpotężniejszych superkomputerów, jest używanych w badaniach meteorologicznych i klimatologicznych [3].

**Z komentarzem [S21]:** Advances in weather prediction

1.Richard B. Alley<sup>1</sup>,

2.Kerry A. Emanuel<sup>2</sup>,

3.Fuqing Zhang<sup>3</sup>

- magazyn Science 25 Jan 2019:

**Z komentarzem [S22]:** Top 500 supercomputers list.  
<http://www.top500.org>

## 2. Cel i zakres pracy

Celem pracy jest stworzenie modelu opartego na technikach głębokiego uczenia oraz wytrenowanie go do prognozowania pogody krótkoterminowej (do 3 dni w przód), a także dogłębna analiza i testy stworzonego modelu. Model powinien zostać dostosowany do jak najlepszej dokładności predykcji wszystkich podstawowych danych synoptycznych.

Praca w ramach projektu została podzielona na 3 etapy, realizowane po kolei:

- wybór i analiza zbioru danych treningowych,
- wybór architektury, budowa i trenowanie modelu,
- testy, analiza wyników i wnioski.

W ramach wyboru zbioru przeszukano zasoby internetowe w celu znalezienia zbioru jak najlepiej odpowiadającego założeniom pracy. Następnie dokonano dogłębnej analizy w celu wykrycia nieprawidłowości i anomalii mogących zaburzyć pracę modelu. Przeanalizowane dane zostały w dalszej kolejności przetworzone do formatu nadającego się do uczenia modelu.

Na etapie budowy modelu, po przeglądzie literatury i istniejących rozwiązań, został stworzony model głębokiej sieci neuronowej produkującej przewidywany stan pogodowy w wybranym oknie czasowym. Zostały także dobrane optymalne parametry oraz zakres przetwarzanych danych dla optymalnej dokładności i wydajności.

Po opracowaniu najlepszego modelu przystąpiono do testów jego skuteczności i użyteczności. Sieć neuronowa z zasady po wytrenowaniu szybko oblicza wyniki, więc wydajność nie podlega testowaniu. Weryfikowano dokładności prognozy w różnych scenariuszach, oraz porównano dokładności prognozy poszczególnych elementów stanu pogodowego. Określono także mocne i słabe strony modelu oraz możliwe poprawki w przyszłości.

## 3. Wprowadzenie do problematyki

W literaturze prognozę pogody dzieli się ze względu na rozmiar okresu prognozowania na [4, 5]:

- „na teraz” – do 2 godzin w przód,
- bardzo krótko-terminowa – do 12 godzin w przód,
- krótkoterminowa - pomiędzy 12 a 72 godziny w przód,
- średnioterminowa - pomiędzy 72 godziny a 10 dni w przód,
- prognoza o zwiększonym zasięgu - pomiędzy 10 dni a 30 dni w przód,
- długoterminowa - ponad 30 dni w przód.

**Z komentarzem [S23]:** <https://www.wmo.int/pages/prog/ww/DPS/GDPS-Supplement5-AppI-4.html>

**Z komentarzem [S24]:** encyklopedia - <http://armchairpatriot.com/Encyclopedias/Encyclopedia%20off%20Weather%20and%20Climate.pdf>

### 3.1. Elementy prognozy stanu pogodowego

W większości prognoz przeznaczenia ogólnego, za czynniki stanu pogodowego uznaje się takie elementy jak:

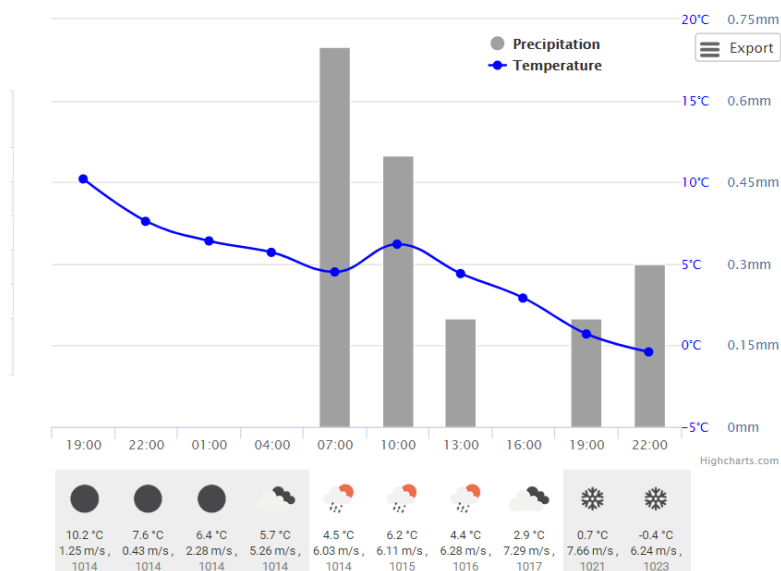
- temperaturę,
- rodzaj i ilość opadów,
- ciśnienie atmosferyczne uśrednione (zazwyczaj znormalizowane do wybranej wysokości),
- wilgotność powietrza,
- zachmurzenie (poziom pokrycia nieba przez chmury, rodzaj chmur),
- zasięg widoczności,
- nasłonecznienie,
- prędkość i kierunek wiatru.

W prognozach rekreacyjnych coraz częściej podaje się prawdopodobieństwo zdarzeń synoptycznych zamiast jednoznacznie je definiować. Na przykład: szansa 30% na mgłę w ciągu następnej godziny.

#### Current weather and forecasts in your city

Main Daily Hourly Chart Map

##### Weather and forecasts in Wrocław, PL



Rys. 3.1 Przykładowa prognoza pogody na 1 dzień z serwisu openweathermap.org

Jak można zauważyć na rysunku 1.2, prognoza nie skupia się na zaprezentowaniu jak największej liczby elementów, a raczej stara się pokazać użytkownikowi w jak najczytelniejszy sposób najistotniejsze informacje. Serwis nie oferuje informacji na temat takich elementów jak widoczność czy kierunek wiatru.

Oprócz prognoz do zastosowań rekreacyjnych istnieją także rozwiązania typowo biznesowe – mające na celu prognozę tylko części parametrów pogodowych dla wąskiej użyteczności w przemyśle, wojsku czy sektorze usługowym. Najważniejszymi zastosowaniami są:

- lotnictwo – zazwyczaj prognozuje się szczegółowo możliwości nagłych zjawisk atmosferycznych – na przykład burz czy gradu, a także poziomu i rodzaju zachmurzenia. Przykładem takiego systemu jest polski system GAMET, monitorujący stan pogody w całym kraju [6].
- marynistyka – prognozy marynistyczne często uwzględniają siłę i kierunek wiatru, a także przewidywane wielkości fal. Kluczowe dla żeglugi jest także wczesne wykrywanie tak zwanych szkwałów czyli nagłych porywów wiatru [5].
- leśnictwo – informacje o wilgotności, na przykład ściółki są powszechnie wykorzystywane do monitorowania zagrożenia pożarowego i pozwalają odpowiednio wcześniej przygotować się do pożaru lasu [5].

### 3.2. Zmienność geograficzna pogody

Ze względu na rozmiar obszaru przewidywań, istnieje co najmniej kilka podziałów prognoz, jeden z nich obejmuje 4 typy:

- prognoza lokalna – obejmuje tylko daną miejscowość lub punkt w miejscu,
- prognoza mezoskalowa – na skali do kilkuset kilometrów,
- prognoza regionalna – dla całych regionów geograficznych,
- prognoza globalna – dla całego globu.

Obecnie zyskują na popularności systemy uniwersalne – mogące po otrzymaniu odpowiednich danych tworzyć prognozy lokalne, dla każdego miejsca na ziemi. Stworzenie takiego systemu jest kosztowne, a same wyniki nie są tak poprawne jak w przypadku modelu stworzonego dla wybranej lokalizacji [7]. Jest tak, ponieważ w różnych strefach klimatycznych inne czynniki (pora roku, godzina) w różnym stopniu wpływają na przyszłą pogodę, odwzorowanie tego przy prognozie jest trudnym i nie zawsze wykonalnym zadaniem.

**Z komentarzem [S25]:** <http://awiacja.imgw.pl/index.php?product=gamet-opis>

**Z komentarzem [S26]:** Encyklopedia - <http://armchairpatriot.com/Encyclopedias/Encyclopedia%20off%20Weather%20and%20Climate.pdf>

**Z komentarzem [S27]:** Encyklopedia - <http://armchairpatriot.com/Encyclopedias/Encyclopedia%20off%20Weather%20and%20Climate.pdf>

**Z komentarzem [S28]:** <https://www.visualcrossing.com/resources/blog/how-do-we-create-our-multi-location-weather-forecast/>



### 3.3. Problem zbierania rozproszonych danych

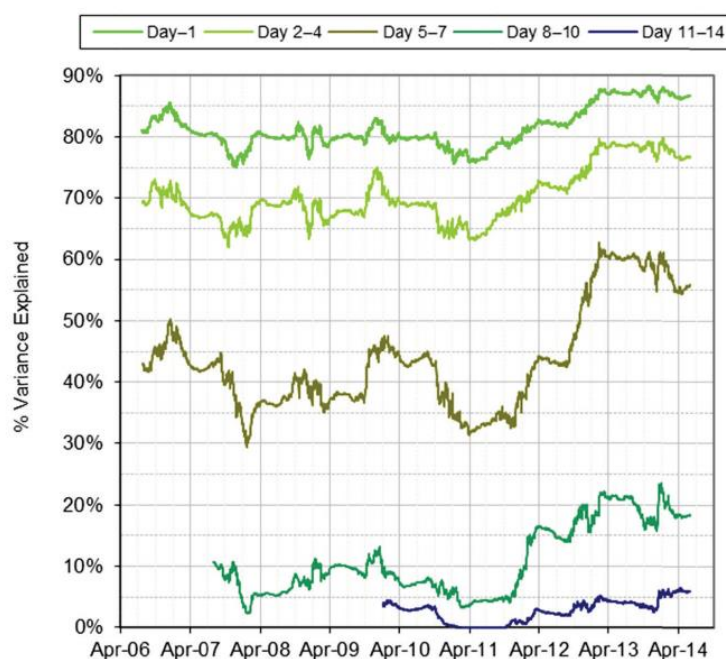
Zbyt mała liczba obserwacji, może znacząco zaniżyć skuteczność prognozy. Dotyczy to zarówno gęstości stacji na badanym obszarze, jak i częstości pobieranych pomiarów. w przypadku prognoz długoterminowych częstość pomiarów nie ma, aż takiego znaczenia, ale dla prognoz „na teraz” brak dokładnych danych może znacząco pogorszyć dokładność prognozy. Kolejnym problemem jest gęstość stacji pomiarowych - aby odpowiednio zamodelować zjawiska zwłaszcza w wielkoskalowych systemach, dane muszą być zbierane wystarczająco gęsto [8]. Utrzymanie dodatkowych stacji pomiarowych generuje dodatkowe koszty, więc ich liczba i częstotliwość pomiarów muszą zawsze być dostosowane do charakterystyki systemu.

**Z komentarzem [S29]:** [https://link.springer.com/chapter/10.1007%2F978-1-935704-20-1\\_3](https://link.springer.com/chapter/10.1007%2F978-1-935704-20-1_3)

### 3.4. Problem malejącej skuteczności prognoz wraz z czasem

Istnieje wiele badań testujących skuteczność poszczególnych modeli, z których większość potwierdza tezę, że skuteczność prognoz drastycznie maleje wraz z czasem. O ile prognozy do 5-dniowe mogą poszczycić się skutecznością nawet do 90%, to prognoza na 14 dni statystycznie nie przekroczy skuteczności 10% [2].

**Z komentarzem [S210]:** <https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.2559>



Rys. 3.2 Porównanie skuteczności prognozy modelu „Climate Forecast System” na przestrzeni czasu dla prognozy 1 dniowej, 2-4 dniowej, 5-7 dniowej, 8-10 dniowej i 11-14 dniowej[2]

Jak można zauważyć, dokładność spada wraz z przewidywanym okresem. Dla prognozy na najbliższy dzień osiąga ponad 90%, a dla prognozy na 14 dzień jest bliska zeru. Można zatem wnioskować, że nawet współcześnie nie jesteśmy w stanie skutecznie przewidzieć pogody w dłuższym okresie czasu.

### 3.5. Metody modelowania pogody

Historycznie wykorzystywano wiele prostych metod na przewidywanie pogody. Na przykład twierdzenie: „Jutro będzie taka sama pogoda jak dzisiaj” sprawdza się w większości przypadków. Od około połowy XIX wieku mówi się o współczesnych metodach synoptycznych. Wraz z wynalezieniem telegrafu (1837 rok) zaczęto zbierać dane z wielu lokalizacji jednocześnie, co pozwoliło ówczesnym meteorologom na sporządzanie tak zwanych map synoptycznych [9].

**Z komentarzem [S211]:** <https://www.britannica.com/science/weather-forecasting/History-of-weather-forecasting>

Modele map synoptycznych były efektywniejsze od znanych wtedy metod, jednak nie były pozbawione wad – z czasem zaczęto pobierać coraz więcej informacji synoptycznych (np. z nowych balonów meteorologicznych) i przetwarzanie liczniejszych informacji w ten sposób stawało się coraz trudniejsze. Rozwiązaniem okazało się powstanie modeli numerycznych. Starają się one odwzorować matematycznie stan atmosfery i jej zmiany, aby zasymulować w jakim stanie znajdzie się za zadany okres czasu. Początkowe modele numeryczne (powstałe krótko po pierwszej wojnie światowej) były obliczane przez człowieka, jednak było to bardzo żmudne zadanie dlatego z czasem zaczęto używać do tego komputerów. Dalszy rozwój meteorologii zawdzięczamy w dużej mierze rosnącej mocy obliczeniowej maszyn. [5,9].

Dzisiaj, modele numeryczne są powszechnie stosowane i uznawane jako najwiarygodniejsze. Nie stosuje się już w żadnym zastosowaniu map meteorologicznych – jako mało wiarygodne, zostały całkowicie wyparte. Rozwiązania współczesne jednak także nie są doskonałe. Najskuteczniejsze systemy, często tworzone przez wiele lat, dalej opierają się na założeniach sprzed kilku dekad. Ich dalszy rozwój jest wyraźnie ograniczony, a stworzenie kompleksowego modelu od podstaw jest czasochłonnym i bardzo kosztownym zadaniem.

W ostatnich latach, coraz częściej wykorzystuje się podejście heurystyczne, oparte na technikach uczenia maszynowego i głębokiego uczenia [10]. Jak pokazują badania, model stworzony w oparciu o syntetyczną sieć neuronową może z podobną dokładnością co modele numeryczne, ale w znacznie krótszym czasie przewidywać stan atmosfery. Obecnie za najskuteczniejsze uważa się modele hybrydowe, używające zarówno metod numerycznych jak i uczenia maszynowego.

**Z komentarzem [S212]:** <https://doi.org/10.1016/j.neune.2006.01.002>

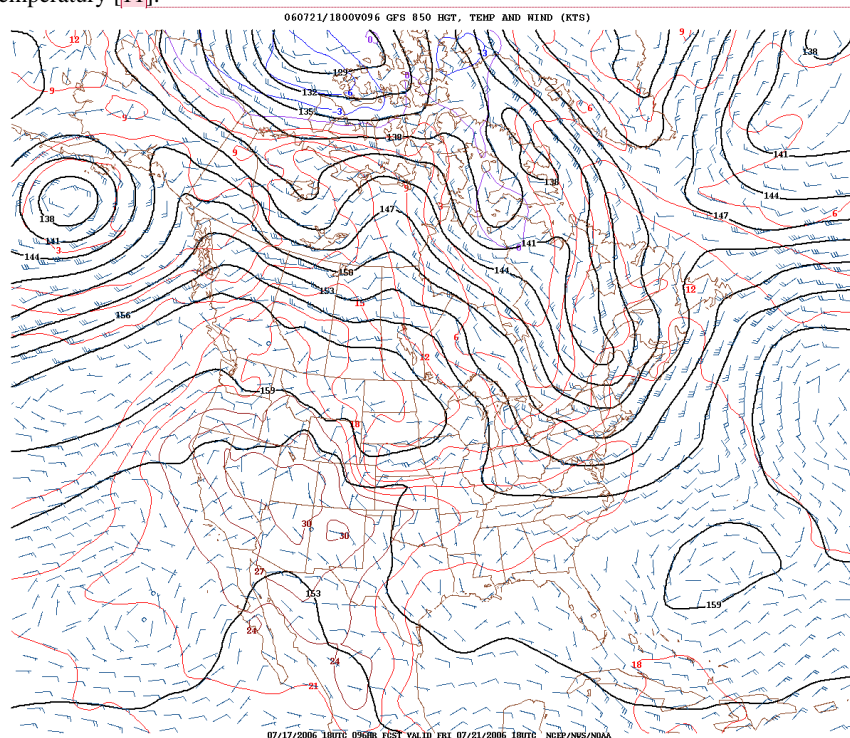
## 4. Przegląd istniejących rozwiązań

### 4.1. Global Forecast System (GFS)

Jest to największy model numeryczny, udostępniany bezpłatnie przez agendę rządową Stanów Zjednoczonych (National Oceanic and Atmospheric Administration). Model uruchamiany jest 4 razy dziennie i generuje prognozę na maksymalnie 16 dni w przód. w prognozie krótko lub bardzo krótkoterminowej informacje są prognozowane w 3-godzinnych interwałach, dla pogody średnioterminowej (powyżej 192 godzin) w interwałach 12-godzinnych.

Obecnie GFS jest uznawany za gorszy w ogólnym przeznaczeniu od konkurencyjnych systemów, jednak najlepiej sprawdzający się do predykcji burz i niebezpiecznych sztormów.

Rys 4.1 przedstawia przykładowy wynik z modelu GFS – prognoza na 96 godzin dla wiatru i temperatury [11].



Rys. 4.1 Przykładowy wynik działania modelu GFS

Z komentarzem [S213]: <https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forecast-system-gfs> - ich strona internetowa

## 4.2. Model European Centre for Medium-Range Weather Forecasts (ECMWF)

Model numeryczny, stworzony i udostępniany bezpłatnie przez Europejskie Centrum Prognoz Średnioterminowych. Generuje głównie prognozę średnio terminową (do 15 dni). Model uruchamiany jest 2 razy dziennie i oprócz prognozy ogólnej prognozuje takie elementy jak: mapę jakości powietrza, dane hydrologiczne, układ atmosfery i inne. Oprócz modelu prognozy średnioterminowej centrum udostępnia także model prognoz 30 dniowych oraz sezonowych do 12 miesięcy w przód. Model ECMWF jest obecnie uważany za najdokładniejszy na świecie dla prognoz średnioterminowych [12].

**Z komentarzem [S214]:** dokumentacja na ich stronie: <https://www.ecmwf.int/en/forecasts/documentation-and-support>

## 4.3. ICOSahedral Nonhydrostatic general circulation model (ICON)

Model powstały dzięki współpracy German Weather Service (DWD) oraz Instytutu meteorologicznego Maxa Plancka. Uważany jest za najskuteczniejszy w prognozach dla Europy. Również jest modelem numerycznym. Prognozuje pogodę na 5 dni w interwałach godzinnych. Predykcja generuje wynik 4 razy na dobę [13,14].

**Z komentarzem [S215]:** <https://www.ecmwf.int/sites/default/files/elibrary/2014/13229-icon-icosahedral-nonhydrostatic-modelling-framework-dwd-and-mpi-m.pdf>

## 4.4. Unified Model(UM)

Model numeryczny, stworzony przez Brytyjski Instytut Meteorologii. Uznawany za najdokładniejszy dla obszaru Wielkiej Brytanii. Generuje prognozę co 12 godzin na 12 godzin w przód. Jest on przewidziany dla działania na obszarze Wielkiej Brytanii, jednak może też skutecznie przewidywać pogodę dla innych regionów i uznawany jako podstawa dla innych regionalnych modeli [15].

**Z komentarzem [S216]:** ich strona internetowa: [https://www.dwd.de/EN/research/weatherforecasting/num\\_modelling/01\\_num\\_weather\\_prediction\\_modells/icon\\_description.html](https://www.dwd.de/EN/research/weatherforecasting/num_modelling/01_num_weather_prediction_modells/icon_description.html)

## 4.5. IBM Deep Thunder

Model hybrydowy – wykorzystujący zarówno elementy numeryczne jak i uczenie maszynowe. Jest to rozwiązanie komercyjne o szerokim zastosowaniu w zależności od potrzeb klienta (np. miasto Dublin wykorzystuje Deep Thunder do ostrzegania przed falami powodziowymi). Model służy głównie do prognoz krótkoterminowych (do 72 godzin) [16,17].

**Z komentarzem [S217]:** <https://www.metoffice.gov.uk/research/approach/modelling-systems/unified-model/weather-forecasting>

## 4.6. Podsumowanie

Jak widać obecnie największą popularnością cieszą się modele numeryczne, dla prognoz średnio i długoterminowych są znacznie bardziej wiarygodne od modeli opartych na uczeniu maszynowym. Prawdopodobnie wynika to z ich rozbudowania. Zaprezentowane modele są tworzone od wielu lat i powstawały w czasach gdy uczenie maszynowe nie było wystarczająco użyteczne, a moc komputerów zbyt mała do stworzenia wiarygodnego rozwiązania. Można, więc postawić hipotezę, że na przestrzeni najbliższych lat się to zmieni i nastąpi dynamiczny rozwój modeli heurystycznych.

**Z komentarzem [S218]:** [http://earthnetworks.com/Portals/0/pdf/IBM\\_Deep\\_Thunder\\_Case\\_Study.pdf](http://earthnetworks.com/Portals/0/pdf/IBM_Deep_Thunder_Case_Study.pdf)

**Z komentarzem [S219]:** Strona projektu: <https://www.ibm.com/ibm/history/ibm100/us/en/icons/deepthunder/>

## 5. Wybór zbioru danych

Specyfiką uczenia maszynowego, jest konieczność znalezienia dostatecznie liczego zbioru danych treningowych. Dane powinny być ustrukturyzowane w kolejne wpisy zawierające kolejne pomiary, spełniające odpowiednie wymagania.

Wybrany zestaw danych, aby nadawał się do prognozy pogody krótkoterminowej, powinien zawierać możliwie jak najczęstsze pomiary. Większość ogólnodostępnych zbiorów zawiera tylko pomiary całonocne, dla przykładu historyczne zapisy pogodowe udostępniane przez Instytut Meteorologii i Gospodarki Wodnej zawierają po jednym pomiarze dziennie z każdej lokalizacji [18].

**Z komentarzem [S220]:** <https://danepubliczne.imgw.pl>

Ponadto, w przypadku pomiarów pobranych z różnych stacji pomiarowych, należy się upewnić, że wszystkie reprezentują jedną strefę klimatyczną – w różnych klimatach istnieją różne zależności pomiędzy poszczególnymi elementami i błędnie dobrany zestaw danych może sprawić, że model nie wyuczy się poprawnie żadnej z nich.

Kolejnym ważnym elementem zbioru jest liczność pomiarów. Zbyt mała liczba próbek nie pozwoli wyuczyć się sieci lub doprowadzi do przeuczenia. Z kolei zbyt duży zbiór nie powinien stanowić problemu, liczne próbki pozwolą uniknąć pogarszającego wyniki oversamplingu. W wypadku skrajnie dużego zbioru możliwe będzie zredukowanie liczby epok uczenia bez pogorszenia skuteczności sieci.

Równie istotne jest także dobranie odpowiednich elementów stanu pogodowego. Zbyt mało elementów w pomiarze – (na przykład sama temperatura) może być niewystarczająca do wyznaczenia zależności pomiędzy nimi i w konsekwencji doprowadzi do sytuacji gdzie model się w ogóle nie uczy. Z kolei zbyt dużo elementów niepotrzebnie wymusza zwiększenie rozmiaru sieci i w konsekwencji powoduje potrzebę użycia sprzętu o większej mocy obliczeniowej. Niepotrzebne elementy pomiarów mogą zostać usunięte w trakcie analizy danych, dlatego na etapie wyboru kierowano się koniecznością znalezienia jak najbardziej szczegółowego zbioru.

### 5.1. Przegląd wybranego zbioru danych

Mając na uwadze przedstawione powyżej zagadnienia, przeszukano dostępne źródła w celu odnalezienia optymalnego zestawu danych treningowych. Rozwiązań spełniających kryteria istnieje wiele, jednak zdecydowano się na pakiet obserwacji meteorologicznych dla obszaru Irlandii, udostępniony przez Met Éireann, czyli narodowy instytut meteorologiczny Irlandii.

Pomiary są dostępne na stronie internetowej instytutu [19], w formie mapy z wyborem parametru i okna czasowego, ale także w gotowych paczkach w formacie CSV. Całość danych jest także dostępna bezpłatnie w serwisie [kaggle.com](https://www.kaggle.com) [20].

**Z komentarzem [S221]:** <https://www.met.ie/climate/available-data/historical-data>

**Z komentarzem [S222]:** <https://www.kaggle.com/connorrot/irish-weather-hourly-data>

Zbiór zawiera w sumie ponad 4 miliony pomiarów, z czego ponad 1,7 miliona wpisów zawiera kompletny zestaw wszystkich elementów. Obserwacje niepełne zostały usunięte, ponieważ modele uczenia maszynowego wymagają spójnego zestawu danych do analizy. Zapisane pomiary odbywały się od początku 1989 roku do końca 2017 roku w 25 różnych stacjach w 15 hrabstwach, rozmieszczonych na całym obszarze Irlandii.

Każdy pomiar składa się z następujących elementów:

- czas wykonania (z dokładnością do pełnej godziny),
- informacje o lokalizacji (nazwa stacji pomiarowej, hrabstwo, współrzędne geograficzne stacji),
- pomiar opadu deszczu (w mm),
- temperatura powietrza (w °C),
- temperatura mokrego termometru - czyli najniższa temperatura, do której przy danej wilgotności i ciśnieniu atmosferycznym można ochłodzić ciało przy pomocy parowania [5] (w °C),
- temperatura punktu rosy - określa wartość temperatury powietrza, do jakiej powietrze musi się ochłodzić, aby osiągnąć stan nasycenia parą wodną (w °C),
- wilgotność względna (w procentach),
- ciśnienie atmosferyczne (w hPa),
- średnia prędkość wiatru (w węzłach),
- dominujący kierunek wiatru (w stopniach),
- czas nasłonecznienia (w godzinach),
- widoczność (w metrach),
- kod synoptyczny – jeden ze 100 zdefiniowanych kodów oznaczających bieżący słowny opis pogody.

Z komentarzem [S223]: encyklopedia



Rys. 5.1 Stacje pomiarowe z największą liczbą pomiarów

Jak można zauważyć, na rysunku 5.1 zaprezentowano położenie 8 najważniejszych stacji pomiarowych które zebrały najwięcej pomiarów. Widoczne jest, że układają się one równomiernie na obszarze wyspy, dzięki czemu odwzorowują dobrze klimat całej Irlandii.

## 5.2. Klimat w wybranej lokalizacji

W Irlandii panuje klimat typu umiarkowanego morskiego wilgotnego. w literaturze opisywany jest duży wpływ mas powietrza znad Oceanu Atlantyckiego oraz wpływ Prądu zatokowego. w obrębie wyspy klimat jest mało zróżnicowany – co jest kluczowe w tym projekcie. Dla przykładu, temperatury zimą wahają się pomiędzy 4 °C a 7 °C, a latem wynoszą od 14 °C do 16 °C. Większość elementów pogodowych jest zbliżona w poszczególnych regionach. Wyjątkiem są zróżnicowane opady, które wahają się od około 750mm rocznie na wschodzie wyspy do nawet 2500mm w górach na zachodnim wybrzeżu.

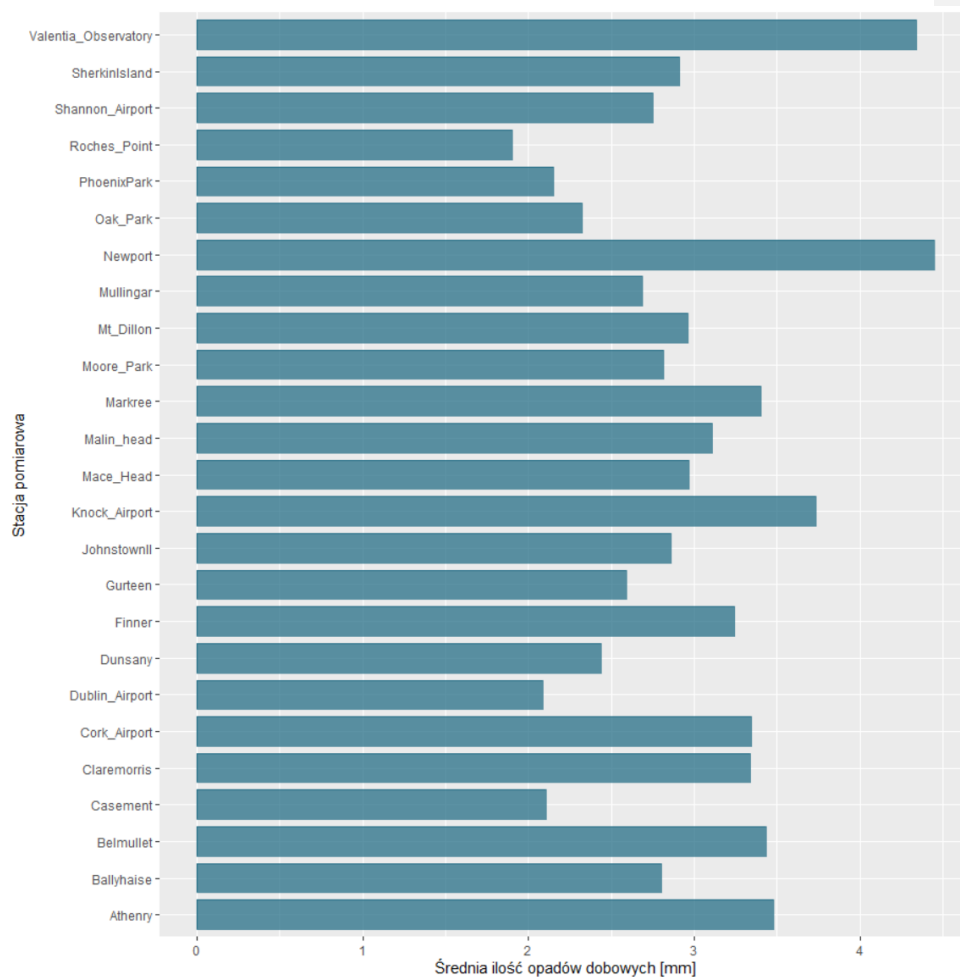
Analizując dane dotyczące zmiany średnich dobowych temperatur na przestrzeni roku w 6 wybranych hrabstwach, wykonano wykres Rys. 5.2.



Rys. 5.2 Zmiany średnich temperatur w wybranych hrabstwach na przestrzeni roku

Jak można zauważyć na rysunku 5.2, różnice temperatur są względnie niewielkie szczególnie latem, kiedy pogoda jest najbardziej zmienna. Istotny jest również fakt, że tendencje zmienności temperatur są takie same we wszystkich regionach na przestrzeni roku, co oznacza że temperatura jest kształtowana przez te same czynniki. w związku z tym, łatwiej jest prognozować zmianę temperatury w czasie.





Rys. 5.3 Średnia dobowa ilość opadów w poszczególnych stacjach pomiarowych

Jak przedstawiono na rysunku 5.3, ilość opadów jest bardzo zróżnicowana, w części stacji zaobserwowano średnio powyżej 4mm dziennie, podczas gdy w punkcie „Roches Point” – na południowym wybrzeżu odnotuje się średnio 1.9 mm opadów na dobę. Tak duże zróżnicowanie może utrudnić funkcjonowanie modelu i w rezultacie pogorszyć wyniki.

Chociaż występujące różnice w opadach znaczące, na tle innych regionów klimat Irlandii wydaje się wyjątkowo stabilny i jednorodny. Oprócz tego, różnice pomiędzy poszczególnymi sezonami są tutaj stosunkowo niewielkie. Oznacza to, wytrenowany model nie powinien mieć zmiennej dokładności w zależności od regionu i sezonu.

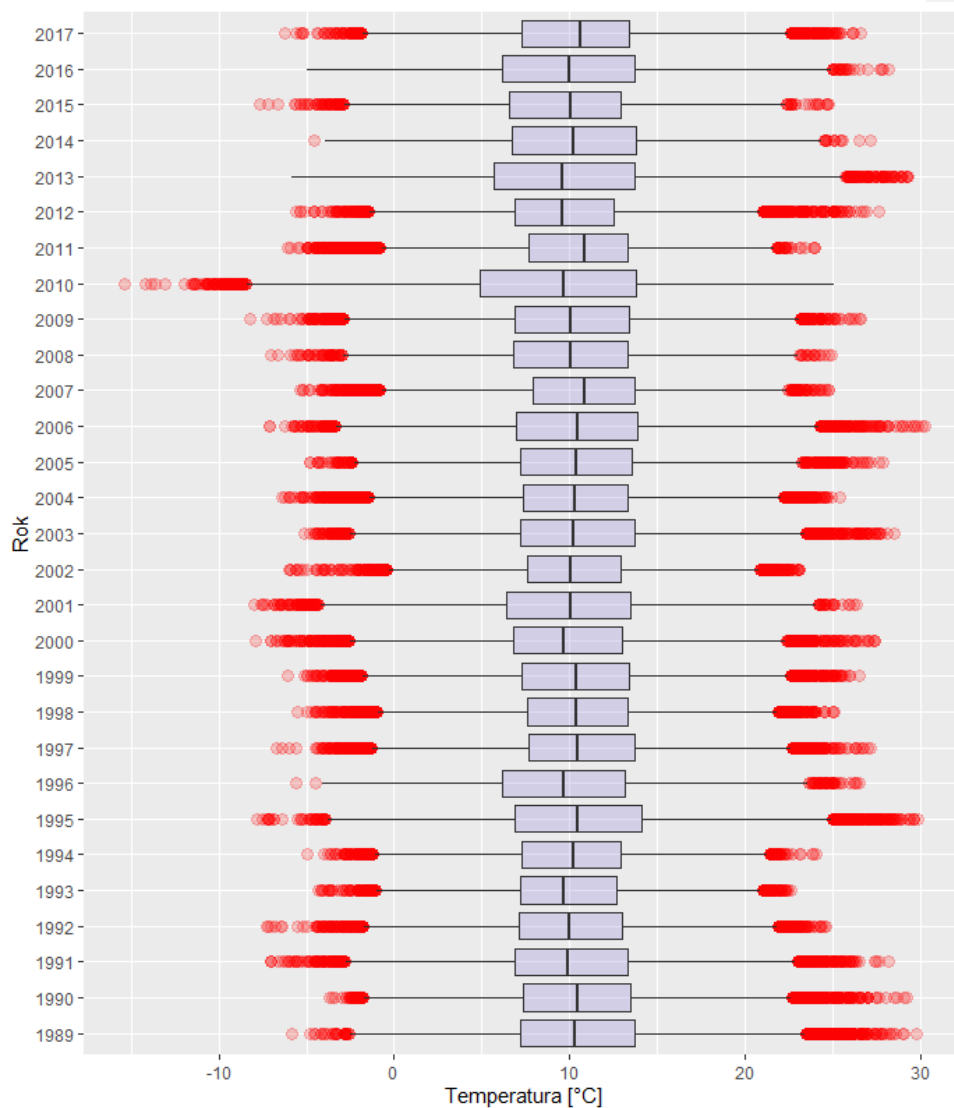
### **5.3. Możliwe anomalie na zbiorze danych**

Wybrany zbiór danych zawiera wystarczająco dużo pomiarów do wytrenowania modelu przy jednorodnym mało zmiennym klimacie Irlandii. Jednak przy tak wielu pomiarach na przestrzeni prawie 40 lat, mogą wystąpić anomalie pogodowe, które powinny zostać wykryte i wyeliminowane, na przykład poprzez usunięcie części pomiarów z danych treningowych. Poniżej zweryfikowano wybrane zjawiska w celu wykrycia pomiarów znacznie odstających od reszty.

#### **5.3.1. Hipoteza: Na skutek zmian klimatu początkowe pomiary znacznie różnią się od końcowych**

W środowisku akademickim temat zmian klimatu jest szeroko omawiany od wielu lat. Większość współczesnych opracowań naukowych skłania się ku twierdzeniu, że klimat zmienia się na całym świecie w coraz większym tempie. Jeżeli na wybranym zestawie danych występują znaczące różnice, część starszych danych może okazać się pogorsząca jakość działania modelu.

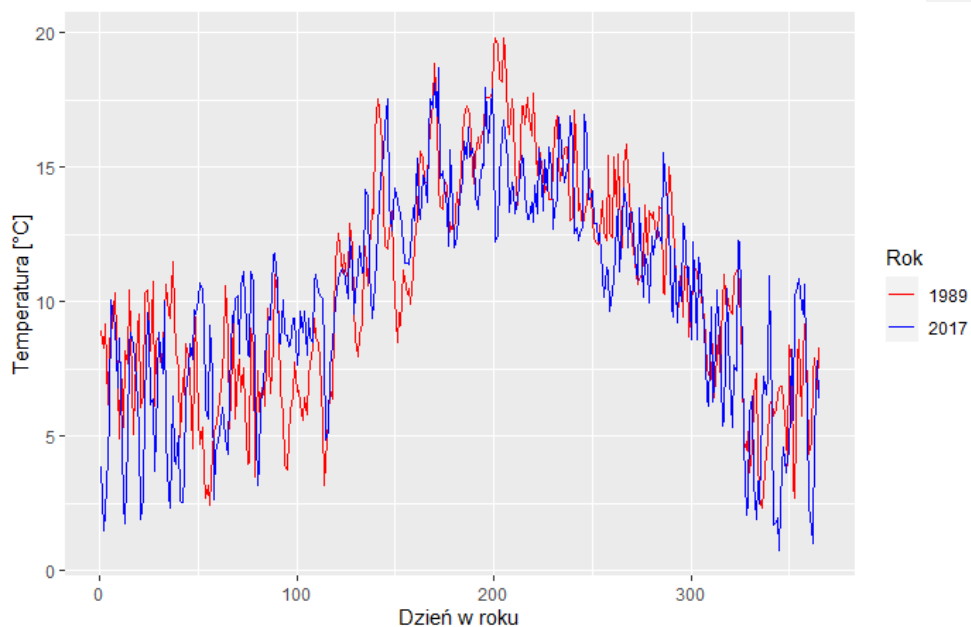
Zmiana klimatu w największej mierze opiera się na zmianie i wzroście temperatury oraz zmianie ilości i gwałtowności opadów, dlatego postanowiono przyrzeć się głównie tym czynnikom.



Rys. 5.4 Rozkład temperatur na przestrzeni lat

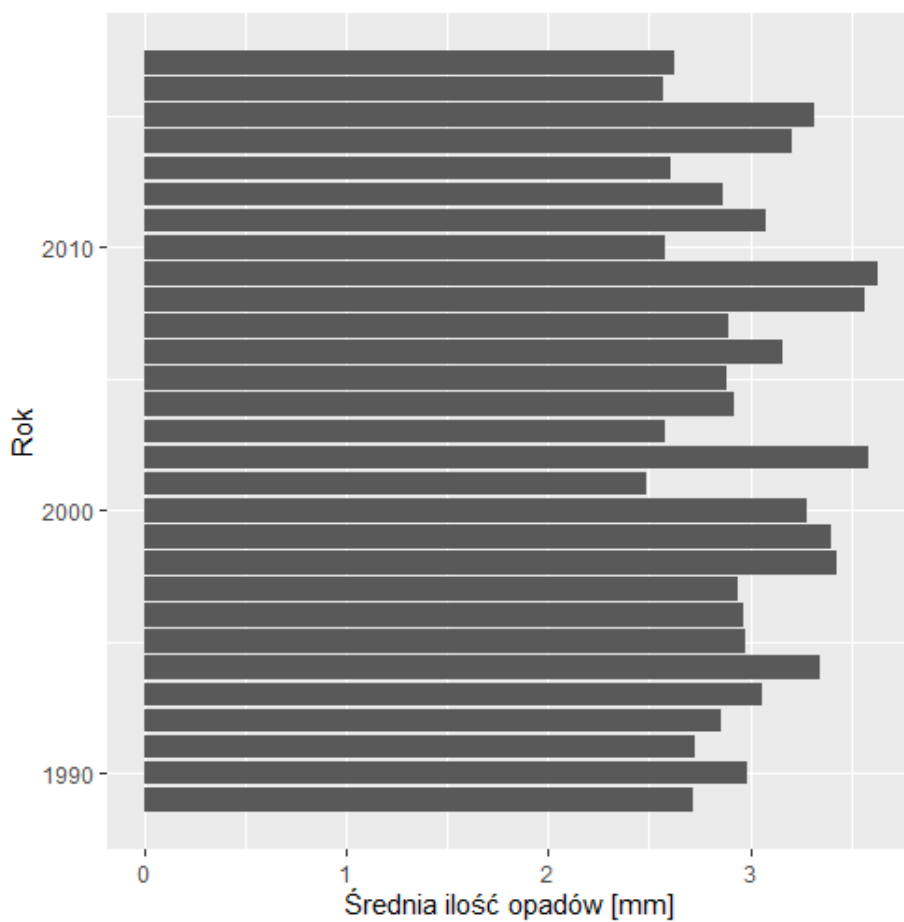
Jak można zauważyć na rysunku 5.4, zarówno średnie temperatury, jak i odstające wartości różnią się wyraźnie w każdym roku pomiarowym, nie mniej jednak nie można zauważyć żadnej tendencji w czasie.

W celu lepszego prześledzenia zmian na przestrzeni pór roku, na wykresie 5.5 zaprezentowano uśrednione temperatury dla roku 1989 i 2017 (skrajnych lat w których wykonywano pomiary).



Rys. 5.5 Porównanie zmian średnich temperatur na przestrzeni całego roku dla lat 1989 i 2017

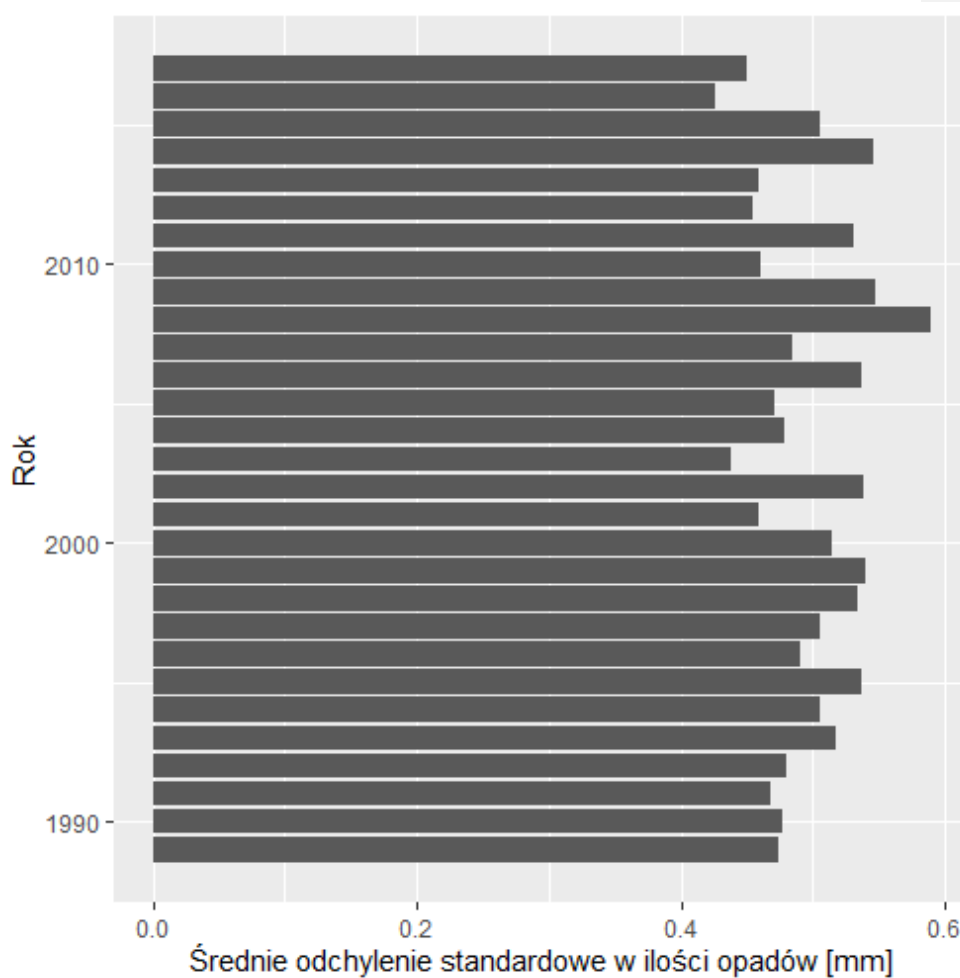
Na wykresie 5.5 nie można zauważyć żadnych znaczących rozbieżności potwierdzających hipotezę. Temperatury różnią się znacząco jednak bez żadnej wyraźnej tendencji. Z powyższych analiz można wywnioskować fakt, iż na badanych danych nie ma anomalii temperaturowych.



Rys. 5.6 Średnia dobowa ilość opadów w poszczególnych latach

Jak można zauważyć na rysunku 5.6, w tym przypadku również nie jest widoczna żadna tendencja. Średnia ilość opadów różni się nawet o około 30% w poszczególnych latach (średnio 2.48mm w 2001 roku i 3.62mm w 2009), jednak nie wykazują żadnego trendu.

Zmiany klimatu często oprócz wpływu na ilość opadów powodują bardziej gwałtowne burze, czy grad. W celu sprawdzenia czy takie odchylenia mają miejsce, na rysunku 5.7 przedstawiono odchylenie standardowe od średnich opadów w poszczególnych latach.



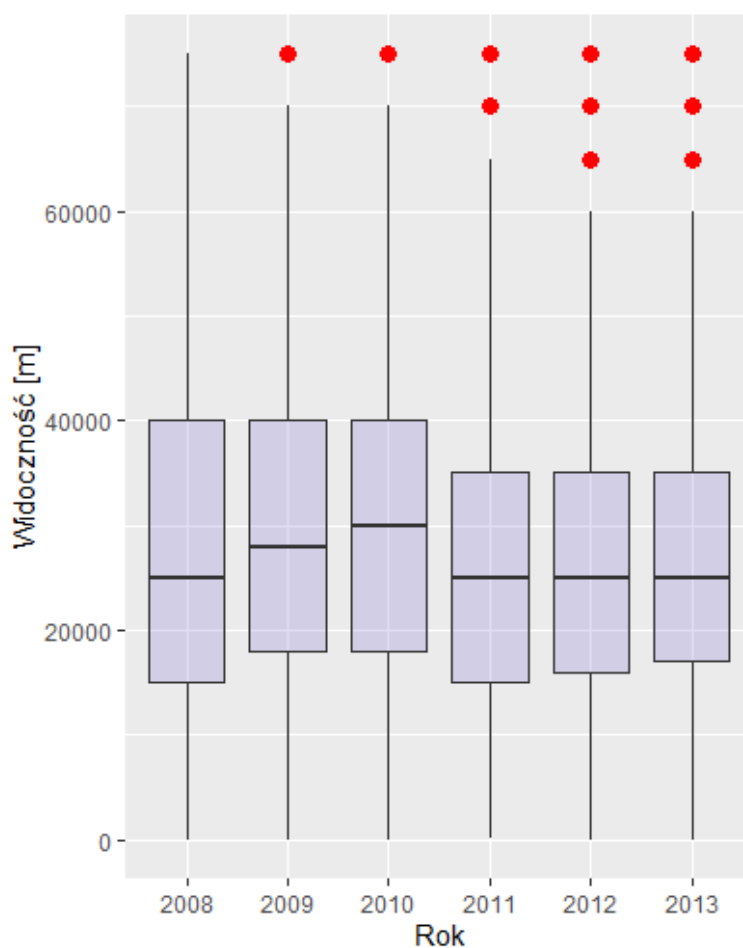
Rys. 5.7 Odchylenie standardowe od średnich dobowych opadów w poszczególnych latach

Jak można zauważyć nie odnotowano wyraźnego zwiększenia gwałtowności opadów. Wartości wahają się od 0.42 mm do 0.58 mm, ale nie wykazują tendencji wzrostowej.

Po przeanalizowaniu powyższych wykresów można wciągnąć wniosek, iż w używanym zbiorze danych nie odnotowano znaczących zmian klimatycznych. Oznacza to, że zarówno pierwsze lata pomiarowe, jak i ostatnie lata są tak samo przydatne w trenowaniu modelu.

### 5.3.2. Hipoteza: Erupcja wulkanu Eyjafjallajökull na Islandii w 2010 roku, spowodowała znaczący spadek widoczności w następnych miesiącach

W kwietniu 2010 roku doszło do serii potężnych wybuchów wulkanu Eyjafjallajökull znajdującego się na Islandii. Po tym zdarzeniu, chmura pyłów utrzymywała się nad Europą przez kilka miesięcy. Zapylenie było na tyle duże, że w większości krajów europejskich zamknięto w związku z tym przestrzeń powietrzną. Chmura pyłów dotarła także nad Irlandię i mogła znacząco wpłynąć na widoczność, oraz w mniejszym stopniu także na inne elementy pogody.



Rys. 5.8 Uśredniona widoczność dla kwietnia, maja i czerwca w latach od 2008 do 2013

Jak można zauważyć na rysunku 5.8, w roku 2010 nie występuje zmniejszenie widoczności, co więcej średnia jest nawet wyższa niż w sąsiednich latach. Można stąd wysnuć wniosek, że dla innych parametrów pogodowych tym bardziej nie występuje żadna anomalia, a erupcja wulkanu nie miała zauważalnego wpływu na stan pogodowy w Irlandii.

#### **5.4. Podsumowanie**

W powyższym rozdziale opisano cechy jakimi kierowano się przy wyborze zbioru danych treningowych, następnie zaprezentowano wybrany pakiet pomiarów meteorologicznych zebranych na obszarze Irlandii. w dalszej części pakiet przebadano pod kątem przydatności do użycia. Analiza wykazała, że pomiary rozkładają się równomiernie i co najważniejsze, reprezentują jedną strefę klimatyczną. Następnie przeprowadzono także testy pod kątem możliwych anomalii zaburzających pomiary w wyniku czego wszystkie hipotezy zostały odrzucone. Można więc wysunąć wniosek że wybrane dane nadają się bardzo dobrze do treningu sieci neuronowej.

### **6. Implementacja modelu**

#### **6.1. Przegląd implementacji występujących w literaturze**

W literaturze opisanych jest wiele rozwiązań wykorzystujących uczenie maszynowe w problemie prognozowania pogody. Wybór modelu powinien być podyktowany przede wszystkim oczekiwanymi wynikami, ale często implementacja jest ściśle zależna od formatu i ilości danych treningowych, a także od dostępnej mocy obliczeniowej.

##### **6.1.1. Sieć neuronowa z jedną warstwą ukrytą (MLP)**

Sieć zwana także perceptronem wielowarstwowym. Jest najprostszym modelem mogącym wykonywać złożoną klasyfikację, stąd też, pierwsze rozwiązania wykorzystujące uczenie maszynowe w meteorologii wykorzystywały perceptron wielowarstwowy.

Model w takiej postaci jest w stanie przyjąć jednorazowo tylko jeden zestaw danych. Stan synoptyczny jest jednak zazwyczaj na tyle skomplikowany, że predykcja na podstawie jednego pomiaru jest niemożliwa – dlatego, aby przekazać dane z wielu próbek pomiarowych przekazuje się do modelu nie konkretne wartości ostatniego pomiaru, a wybrane wartości statystyczne z wielu ostatnich. Na przykład, zamiast podawać ostatnią zmierzoną temperaturę w celu przewidzenia następnej, podaje się: średnią, medianę minimum, maksimum i odchylenie standardowe z ostatnich 100 pomiarów.



Główną cechą rozwiązań z jedną warstwą ukrytą jest, w stosunku do innych modeli prosta budowa, a co za tym idzie mała liczba parametrów. Najważniejszą zaletą jest stosunkowo szybkie uczenie – co w przypadku systemów o ograniczonych zasobach może okazać się kluczowe. Jednak pojemność modelu jest zależna przede wszystkim od rozmiaru warstwy ukrytej i przy rozsądnych wartościach jest stosunkowo niewielka. Zatem główną wadą perceptronu wielowarstwowego jest mała pojemność informacyjna, która może uniemożliwić poprawną predykcję przy skomplikowanych zależnościach meteorologicznych. [21, 22]

### 6.1.2. Sieć rekurencyjna (RNN)

Ideą sieci rekurencyjnej jest wykorzystywanie sekwencji informacji (w odróżnieniu od tradycyjnych MLP gdzie wejścia są z założenia niezależne od siebie). Wynik sieci rekurencyjnej zależy nie tylko od zadanych danych wejściowych ale również od danych z poprzedniego wejścia (lub wyjścia) sieci. RNN zazwyczaj są stosowane do predykcji kolejnych wyrazów ciągów, takich jak kolejne słowa w problemie rozpoznawania mowy, czy kolejne stany atmosfery w prognozie pogody [23,24].

Sieci rekurencyjne mogą dobrze wychwytywać zależności pomiędzy zmianami pogody i nauczyć się je odtwarzać – pod tym względem mogą okazać się znacznie skuteczniejsze od MLP. Jednak podzielenie danych na poprawne podciągi jest czasochłonne, a uczenie sieci na całych ciągach liczb może być dużo wolniejsze i wymagać wydajniejszego sprzętu.

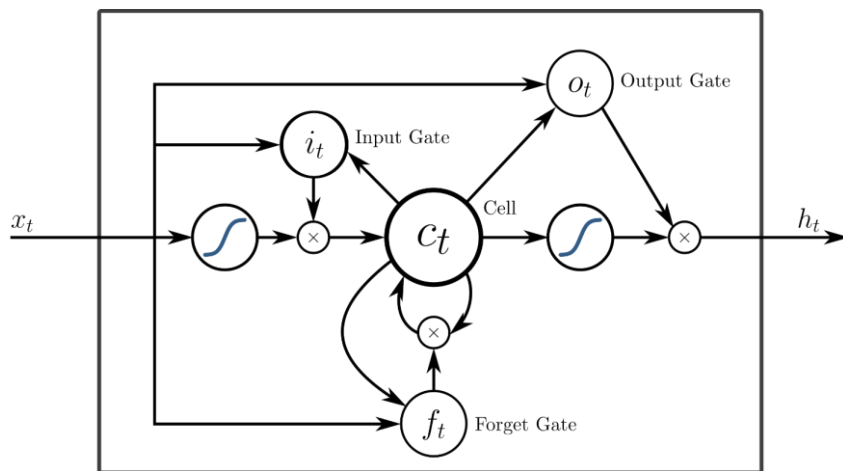
Szczególnym przypadkiem sieci rekurencyjnej jest LSTM (Long Short-Term Memory). Sieć LSTM rozwiązuje problem zaniku informacji odleglejszych w czasie niż jedna iteracja. Pozwala to na dłuższe przechowywanie informacji które w klasycznym RNN zostałyby już zapomniane. w prognozowaniu może być to kluczowe na przykład w przypadku, gdy próbujemy przewidzieć, kiedy obecny deszcz się skończy – do tego może być przydatna informacja jakie było ciśnienie i jaki rodzaj zachmurzenia przed rozpoczęciem samych opadów.

**Z komentarzem [BK(24)]:** <https://reader.elsevier.com/reader/sd/pii/S221201731200326X?token=E29DDC2447AE28E7D7C4C583FABE80A5B2B09AF81E4C8B66D64CF04D7BDC291CFE16C164C7C37B2D3F60AF1F69C61DE0>

**Z komentarzem [BK(25)]:** Imran Maqsood Muhammad Riaz Khan Ajith Abraham An ensemble of neural networks for weather forecasting

**Z komentarzem [BK(26)]:** <https://ieeexplore.ieee.org/abstract/document/7508408>

**Z komentarzem [BK(27)]:** <https://ieeexplore.ieee.org/abstract/document/7415154>



Rys. 6.1 Schemat komórki LSTM

Na rysunku 6.1 przedstawiono schemat budowy komórki sieci LSTM. Jak można zauważyć składa się ona z 3 bramek: wejściowej, wyjściowej oraz bramki zapominania. Komórka przechowuje informacje w stanie długoterminowym, przechodząc przez bramkę zapominania, modyfikuje informacje tam zawarte. w procesie uczenia sieci, komórka sama uczy się jakie informacje przechowywać a jakie zapominać w pamięci długoterminowej.

### 6.1.3. Sieć konwolucyjna (CNN)

Model głębokiej sieci, służący głównie do przetwarzania obrazów. Warstwy konwolucyjne sieci mogą skutecznie wychwytywać cechy z dużej ilości złożonych danych wejściowych, dla przykładu wykrywać specyficzne kształty na obrazach. Im więcej model zawiera warstw konwolucyjnych tym bardziej skomplikowane cechy jest w stanie rozpoznać. [24,25]

Sieci konwolucyjne są bardzo przydatne przy analizie i klasyfikacji danych o wysokiej rozdzielczości, ale w odróżnieniu od sieci rekursyjnych nie mają możliwości zapamiętywania informacji historycznych. Dodatkowo, charakteryzują się bardzo długim czasem uczenia i wymaganiem dużej mocy obliczeniowej. w kontekście prognozy pogody jest to często używane do analizy i klasyfikacji zdjęć satelitarnych lub map synoptycznych.

### 6.1.4. Sieć konwolucyjna LSTM

Rozwiązanie hybrydowe, łączące cechy zarówno sieci konwolucyjnej jak i LSTM. Wykorzystywane zazwyczaj w prognozie pogody na podstawie danych z radarów i satelitów. Pozwala na przetwarzanie bardzo dużej ilości danych synoptycznych, jednak obecne rozwiązania nie są jeszcze dopracowane, dlatego skuteczność może bardzo mocno zależeć od przypadku użycia. w literaturze istnieje zaprezentowanych wiele modeli tego typu, jednak zazwyczaj ich budowa jest ściśle powiązana z konkretnym przypadkiem. [26,27]

**Z komentarzem [S228]:** autokorekta nie zna słowa konwolucja?

**Z komentarzem [BK(29)]:** <https://arxiv.org/ftp/arxiv/papers/1506/1506.01195.pdf>

**Z komentarzem [BK(30)]:** <https://www.sciencedirect.com/science/article/abs/pii/S0925231218311214>

**Z komentarzem [BK(31)]:** <http://papers.nips.cc/paper/5955-convolutional-lstm-network-a-machine-learning-approach-for-precipitation-nowcasting>

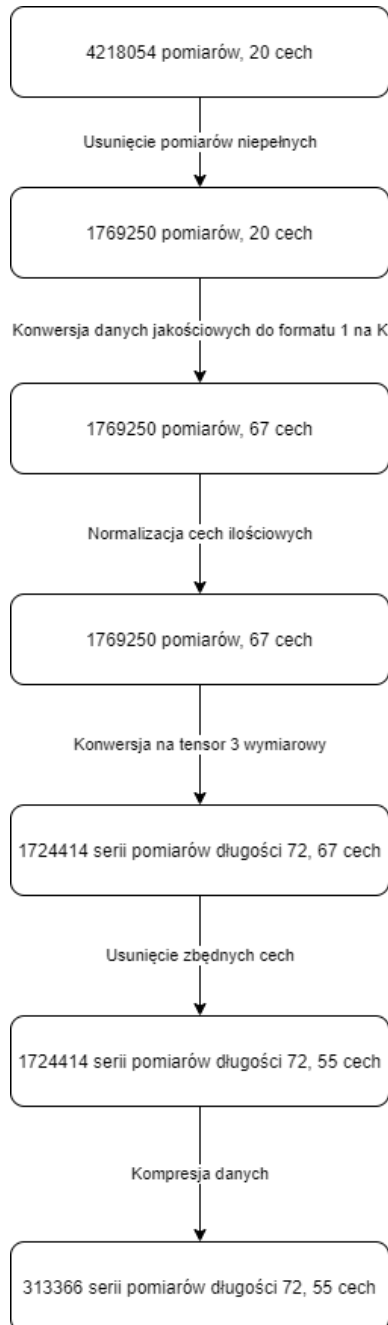
### **6.1.5. Podsumowanie**

Powyżej przedstawiono najczęstsze implementacje dotyczące problemu prognozy pogody. Istnieje także wiele innych, bardziej niszowych rozwiązań, jednak ze względu na dobre udokumentowanie w literaturze, zdecydowano się rozważać jedynie te wymienione.

Po rozważeniu wszystkich rozwiązań, biorąc po uwagę zarówno cele projektu, wybrany zbiór danych, jak i ograniczenia sprzętowe zdecydowano się na implementację wielowarstwowej sieci z warstwą LSTM. Ta metoda pozwala na osiągnięcie bardzo dobrych wyników, a wybrany zbiór danych jest odpowiedni do podziału na serie czasowe. Zrezygnowano z użycia sieci konwolucyjnych głównie ze względu na małą ilość parametrów danych wejściowych – kilkanaście ilościowych i maksymalnie kilka jakościowych to zbyt mało na skuteczne wykorzystanie konwolucji.

### **6.2. Przetworzenie danych treningowych**

Wybrany zbiór pomiarów musi zostać odpowiednio przetworzony, aby można było użyć go do treningu sieci. Między innymi oznacza to oczyszczenie próbek ze zbędnych informacji a także podzielenie na ciągi czasowe potrzebne do warstwy rekurencyjnej. Na rysunku 6.2 przedstawiono uproszczony schemat przetworzenia danych.



Rys. 6.2 Algorytm przetwarzani danych treningowych

Jak można zauważyć na rysunku 6.2, dane zostały przetworzone sekwencyjnie do stanu nadającego się do uczenia sieci neuronowej. Można zauważyć, że z pierwotnych ponad 4 milionów rekordów powstało około 300 tysięcy ciągłych serii. Poszczególne etapy przetwarzania zostaną omówione poniżej.

#### **6.2.1. Usunięcie danych niepełnych**

Spośród wszystkich zebranych pomiarów, ponad połowa okazała się nie zawierać wszystkich cech. Najczęściej nie mierzono widoczności (2,44 miliona) oraz poziomu nasłonecznienia (2,41 miliona). Dane treningowe nie mogą zawierać pustych informacji, dlatego w przypadku braku chociaż jednej cechy usuwano cały pomiar.

#### **6.2.2. Konwersja do formatu 1 na K (One hot encoding)**

Składowe pomiaru, które nie są cechami ilościowymi a jakościowymi, nie mogą być bezpośrednio przetworzone przez sieć. W tym celu należy z każdej z nich utworzyć listę cech ilościowych zawierających wszystkie kategorie i przyjmujących wartości jedynie 0 lub 1. Potocznie nazywa się to notacją 1 z K, lub kodem 1 z n. W analizowanym przypadku taka sytuacja występuje dla kodu synoptycznego oraz różnych parametrów czasu wykonania pomiaru.

Kod synoptyczny oznaczony jest numerem od 0 do 99 i reprezentuje opis słowny bieżącego stanu pogodowego. Dla przykładu, kod 81 oznacza: „średni lub duży opad deszczu”. Dodatkowo, kody zostały tak rozplanowane aby można je było uprościć do 10 mniej szczegółowych kodów. Implementacja rozpoznawania 100 kodów wiązałaby się z nieproporcjonalnym zwiększeniem wektora cech wejściowych, dlatego zdecydowano się na użycie uproszczonych kodów synoptycznych. Tak więc do pomiarów wejściowych dodano 10 cech z wartościami 0 i 1 oznaczającymi przynależność pomiaru do danej klasy.

Czas wykonania pomiaru także jest cechą jakościową. Jednak zmienność pogody w czasie jest bardziej skomplikowana, ponieważ stan pogodowy jest zmienny w różnych okresach czasowych. Wyróżniono 3 możliwe zależności pogody w okresach czasowych: zależność od roku pomiarowego, od pory roku i od pory dnia. Hipoteza, że istnieją znaczące różnice pomiędzy latami została odrzucona, zatem wydzielono tylko cechy odpowiedzialne za porę roku i porę dnia.

W kwestii pory roku zdecydowano podzielić rok na 12 miesięcznych okresów, tak aby odpowiednio uwzględnić różnice pomiędzy rozmytymi sezonami. Podobnie jak w przypadku kodów synoptycznych, wydzielono 12 dodatkowych cech w wektorze wejściowym i oznaczono przynależność do danego miesiąca.

W podobny sposób postanowiono wydzielić cechy odpowiadające porze dnia. Pomiary są zapisane z dokładnością do pełnej godziny, więc aby jak najlepiej oddać ważną informację godzinową zdecydowano się na wydzielenie kolejnych 24 elementów wektora w celu zakodowania godziny pomiaru – analogicznie do pór roku.

### 6.2.3. Normalizacja cech ilościowych

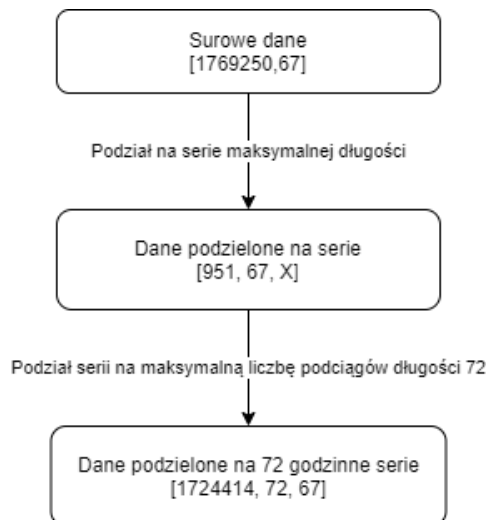
Zebrane pomiary występują w różnych jednostkach i osiągają różne rzędy wartości. Dla przykładu – widoczność przyjmuje wartości od 0 do 10000m. Przekazanie skrajnie różnych wartości do modelu może uniemożliwić jego poprawne uczenie. Dlatego znormalizowano wszystkie zmienne ilościowe za pomocą standaryzacji, według wzoru:

$$x' = \frac{x - \mu}{\sigma} \quad (6.1)$$

Gdzie  $\mu$  to średnia, a  $\sigma$  to odchylenie standardowe. Cechy jakościowe nie wymagają skalowania, ponieważ po przetworzeniu w poprzednim etapie wszystkie elementy przyjmują tylko wartości 0 i 1. Ustandaryzowane pomiary nie mają zamkniętych granic wartości, ale tylko wyjątkowo odstające wartości wychodzą z zakresu  $[-3, 3]$ .

### 6.2.4. Konwersja na tensor 3 wymiarowy

Ideą sieci rekurencyjnej jest wychwytywanie zależności pomiędzy kolejnymi wyrazami ciągu. w procesie uczenia sieci należy podawać kolejne wyrazy ciągów lub jeden kompletny ciąg. Całego zbioru nie można traktować jak jeden kompletny ciąg, ponieważ nie są to kolejne pomiary z jednej stacji. Pomiary były zbierane na 25 stacjach, do tego nie są one ciągłe – często występują przerwania i nieciągłości. Dlatego, z macierzy danych wydzielono trzeci wymiar w którym zapisano kolejne następujące po sobie próbki. Dodatkowo, w wybranej technologii implementacji modelu (Python, Keras) serie muszą mieć jednakową długość.



Rys. 6.3 Uproszczony schemat algorytmu konwersji danych na serie

Jak można zauważyć na rysunku 6.3, w pierwszym kroku zbiór jest dzielony na maksymalnej długości podciągi nieprzerwanych pomiarów w jednej lokalizacji. w tym wypadku serii jest 951, każda jest innej długości.

W drugim kroku, serie są sprowadzane do wspólnej długości – 72. Prognozę krótkoterminową definiujemy jako predykcję maksymalnie 3 dni w przód, więc nie jest potrzebne utrzymywanie dłuższych serii – 72 kolejne pomiary różniące się w czasie o godzinę, tworzą 3 dniowy ciąg. Aby to zrobić dla każdego elementu ciągu jest tworzony nowy ciąg składający się z niego i 71 kolejnych elementów. Jeżeli za elementem nie ma wystarczającej liczby kolejnych, algorytm przechodzi do kolejnej serii. Dla przykładu, jeżeli istnieje 100 elementów w serii, to zostanie stworzonych 28 serii długości 72. Wynikiem działania algorytmu jest tensor zawierający dane podzielone serie o równej długości o kształcie [1724414, 72, 67].

### 6.2.5. Kompresja danych

Tensor powstały w wyniku operacji w poprzednim kroku, zawiera prawie 5 miliardów liczb zmiennoprzecinkowych. w kontekście uczenia samej sieci nie jest to zbyt dużo – nie ma górnego ograniczenia na ilość danych. Jednak tak duża struktura danych wymaga kosztownej konfiguracji sprzętowej. Ponieważ nie przewiduje się uruchamiania modelu na systemie z wieloma kartami graficznymi o dużej pojemności pamięci, dane zostały odpowiednio zredukowane.

Pierwszym krokiem jest konwersja wszystkich liczb zmiennoprzecinkowych do formatu float32, czyli 2-krotnie mniejszego niż standardowy float64.

Następnie, przystąpiono do redukcji serii danych, zauważając że każde dwie serie, różnią się tylko jednym pomiarem godzinowym – czyli w 71/72 stanowią tę samą informację. Jeżeli zatem usunąć co drugą serię, w dalszym ciągu sieć zachowałaby 71/72 informacji początkowej. Biorąc pod uwagę możliwe straty w zawartych informacjach, oraz jak największe ograniczenie rozmiaru zbioru danych, zdecydowano się usunąć 80% serii, czyli pozostawić 1 na 5 kolejnych serii. w ten sposób, tensor zachowuje ponad 93% pierwotnych informacji, zmniejszając swój rozmiar pięciokrotnie.

### 6.2.6. Usunięcie zbędnych cech

Na tym etapie usunięto elementy pomiaru, które nie są powiązane z pogodą. Jak w poprzednim dziale już wykazano położenie stacji pomiarowej nie ma znaczącego wpływu na pomiary, więc wszystkie elementy definiujące lokalizację są zupełnie zbędne – m.in. informacje o hrabstwie, czy nazwa stacji. Usunięto także kod synoptyczny oraz datę pomiaru, ponieważ informacje wydobyte już wcześniej są wystarczające.

### 6.3. Opracowanie struktury sieci

W literaturze opisane są liczne badania, próbujące zgłębić temat wykorzystania sieci rekurencyjnych do prognozowania pogody. Pomijając rozwiązania bardziej złożone – wykorzystujące również warstwy konwolucyjne, najczęściej spotykane są model oparte na:

- A. pojedynczej warstwie LSTM lub RNN i jednej lub dwóch warstwach ukrytych w pełni połączonych [28],
- B. wielu (od 2 do 4) kolejnych warstwach LSTM i co najwyżej jednej warstwie ukrytej w pełni połączonej [29, 30].

Wszystkie znalezione rozwiązania były testowane na innych danych, zatem nie można wprost określić, która architektura jest skuteczniejsza. Podejście A może bardziej sprawdzić się w określaniu kluczowego parametru – w literaturze często występuje w kontekście przewidywania opadów deszczu. Może się to okazać mało skuteczne w rozważanym problemie. Jednak z drugiej strony, sieć zbudowana w ten sposób powinna być znacząco mniej wymagająca pod względem mocy obliczeniowej. Każda kolejna warstwa wymaga wykonania coraz większej liczby operacji, a warstwy LSTM są pod tym względem bardziej wymagające od klasycznych warstw w pełni połączonych. Uczenie sieci w architekturze B, prawdopodobnie zajmie znacząco więcej czasu lub może być w ogóle niewykonalne na słabszych maszynach.

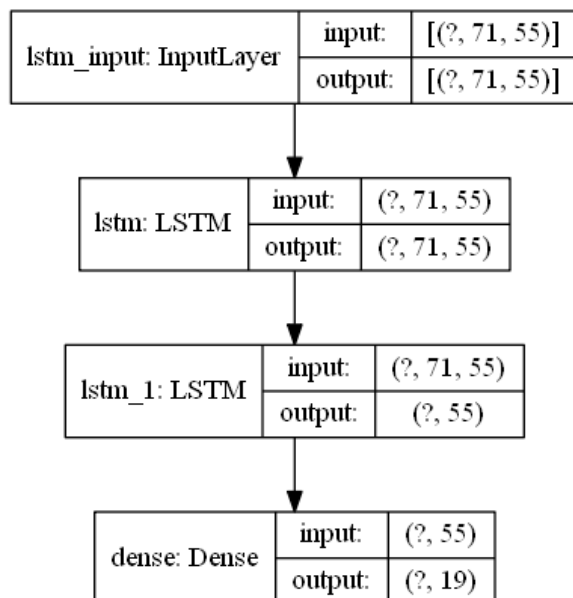
Biorąc pod uwagę przeznaczenie sieci, ilość dostępnych danych treningowych, a także możliwości sprzętowe platformy testowej, zdecydowano się na użycie sieci z 2 warstwami LSTM oraz jedną warstwą ukrytą w pełni połączoną.

**Z komentarzem [BK(32)]:** <https://doi.org/10.3390/w11071387>

**Z komentarzem [BK(33)]:** [https://www.researchgate.net/profile/Mohamed\\_Akram7/publication/304066008\\_Sequence\\_to\\_Sequence\\_Weather\\_Forecasting\\_with\\_Long\\_Short-Term\\_Memory\\_Recurrent\\_Neural\\_Networks/links/58aecee345851503be945379/Sequence-to-Sequence-Weather-Forecasting-with-Long-Short-Term-Memory-Recurrent-Neural-Networks.pdf](https://www.researchgate.net/profile/Mohamed_Akram7/publication/304066008_Sequence_to_Sequence_Weather_Forecasting_with_Long_Short-Term_Memory_Recurrent_Neural_Networks/links/58aecee345851503be945379/Sequence-to-Sequence-Weather-Forecasting-with-Long-Short-Term-Memory-Recurrent-Neural-Networks.pdf)

**Z komentarzem [BK(34)]:** <https://doi.org/10.1016/j.procs.2018.08.153>





Rys. 6.4 Schemat budowy sieci

Jak można zauważyć na rysunku 6.4, sieć składa się z 2 warstw LSTM, gdzie obie zawierają tyle samo neuronów. Domyślnie jest to 71, ale te wartości będą szczegółowo dobierane na etapie doboru hiperparametrów. Następnie znajduje się warstwa w pełni połączona, która pełni funkcję warstwy wyjściowej. Sieć przyjmuje kolejne wyrazy ciągu 55 elementowe, ale zwraca tylko 19, ponieważ nie ma potrzeby obliczania wartości czasu pomiaru, tak jak na wejściu, a jedynie stan pogody, dlatego ostatnia warstwa składa się z 19 neuronów.

#### 6.4. Funkcja straty

Jednym z najważniejszych parametrów sieci jest tak zwana funkcja straty, zwana także funkcją kosztu. Koszt jest to w uproszczeniu odległość pomiędzy wektorem docelowym, a wektorem wygenerowanym przez sieć, jest on używany do obliczania gradientu, a później do korekcji wag sieci. Koszt może także służyć do ogólnej oceny skuteczności sieci. Najczęściej wykorzystuje się funkcje sumujące odległości pomiędzy poszczególnymi cechami i wartościami ich predykcji przez sieć lub kwadraty tych odległości. Szczegółowy wybór funkcji straty zostanie omówiony w dalszej części pracy.

W tym wypadku funkcja kosztu jest wyliczana nie z odległości pomiędzy całym wektorem wynikowym a prawdziwym, a tylko z jego 19 pierwszych elementów czyli wszystkich nie określających czasu pomiaru. Jak można zauważyć, sieć zwraca tylko te 19 elementów.

## 6.5. Uczenie sieci

W procesie uczenia sieci jest prezentowane 71 próbek, następnie obliczana jest różnica pomiędzy przewidzianą 72 próbką, a rzeczywistym ostatnim pomiarem w serii. Różnica obliczana za pomocą funkcji kosztu jest następnie propagowana na kolejne warstwy jako korekta wag. Jest to tak zwana propagacja wsteczna – obecnie jedna z najczęściej stosowanych metod korekcji wag [31].

Do doboru korekcji wag, czyli optymalizacji sieci, najczęściej wykorzystuje się algorytm stochastycznego spadku wzdłuż gradientu lub jego modyfikacje [32]. Wybór algorytmu i porównanie ich zostaną przedstawione w dalszej części pracy.

## 6.6. Funkcje aktywacji kolejnych warstw

Pobudzenie każdego neuronu jest zależne od sumy pobudzenia neuronów z warstwy poprzedniej z nim połączonych. Zależność pomiędzy tymi wartościami może być liniowa, jednak zazwyczaj stosuje się do tego odpowiednie funkcje w większości o zamkniętej wartościowości z przedziału  $[-1, 1]$  lub  $[0, 1]$ .

Skoro dane wejściowe są ustandaryzowane, to mogą osiągać wartości spoza tego zakresu, dlatego w warstwie wyjściowej zdecydowano się nie stosować funkcji o zamkniętym przedziale wartości, a użyć funkcji liniowej.

W dalszej części pracy, zaprezentowano dobór funkcji aktywacji dla dwóch wewnętrznych warstw LSTM.

## 7. Wybór hiperparametrów modelu

Wynik uczenia sieci w dużej mierze zależy od wyboru odpowiedniej architektury, ale również od odpowiedniego doboru parametrów. Ponieważ wartości optymalnych atrybutów są bardzo zależne od konkretnych danych, nie można dobrać ich poprzez analogie względem podobnych projektów. W tym rozdziale zostaną przedstawione testy kluczowych hiperparametrów oraz wybrane ich optymalne wartości.

### 7.1. Platforma testowa

Model uczono na komputerze z systemem Windows 10, posiadającym 12 rdzeniowy procesor Intel i7 8700K, 16GB pamięci RAM, oraz kartę graficzną NVIDIA GeForce 2070 RTX z 8GB pamięci VRAM. Sieć została zaimplementowana z języku Python, przy użyciu bibliotek Keras oraz Tensorflow. Użyto technologii Nvidia CUDA do trenowania modelu na karcie graficznej.

**Z komentarzem [BK(35)]:** Książka: Hands-On Machine Learning with Scikit-Learn and TensorFlow. Concepts, Tools, and Techniques to Build Intelligent Systems

Rozdział: Multi-Layer Perceptron and Backpropagation

**Z komentarzem [BK(36)]:** <https://ruder.io/optimizing-gradient-descent/index.html#whichtochoosetouse>

**Z komentarzem [BK(37)]:** Książka: Hands-On Machine Learning with Scikit-Learn and TensorFlow. Concepts, Tools, and Techniques to Build Intelligent Systems

Rozdział: Faster Optimizers

## 7.2. Metoda testowania i ograniczenia

Użytkowana platforma testowa dysponuje stosunkowo dużą mocą obliczeniową, mimo to niektóre przypadki testowe wykonywały się wiele godzin. Jako wynik testu uznano błąd na zbiorze walidacyjnym, stanowiącym 1/3 całego zbioru. Model w każdym przypadku uczono do momentu, aż błąd nie poprawił się przez co najmniej 10 epok lub czas uczenia przekroczył 3 pełne godziny.

## 7.3. Wybór algorytmu optymalizacji sieci

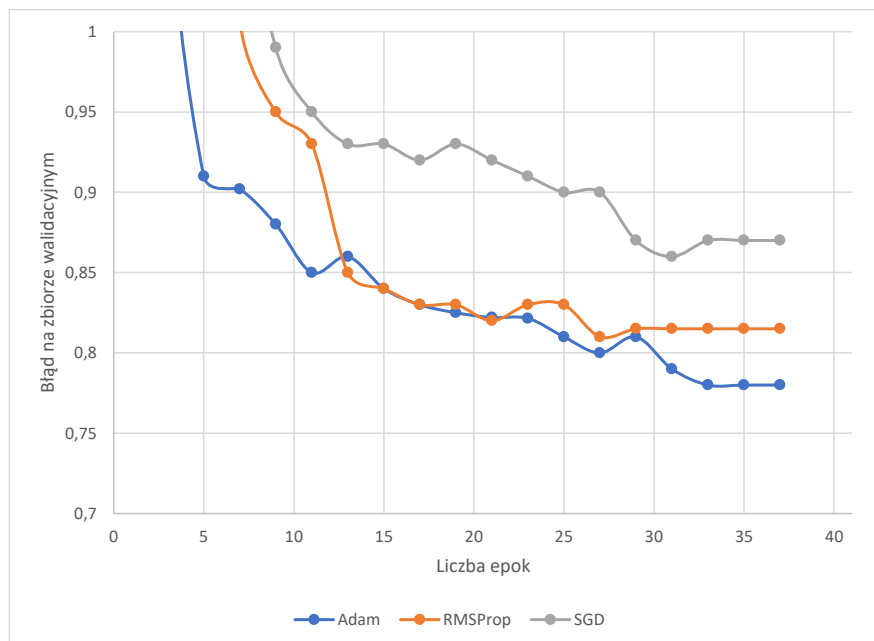
W uczeniu przez propagację wsteczną najstarszym i najlepiej opisanym algorytmem jest spadek gradientu stochastycznego (SGD), jednak w ostatnich latach, coraz więcej popularności zyskują nowsze, bardziej złożone algorytmy. Poniżej zaprezentowano wyniki testów dla 3 algorytmów:

- Spadek gradientu stochastycznego – podstawowy algorytm minimalizacji funkcji kosztu. Opiera się na znalezieniu spadku funkcji straty i podążaniu zgodnie z nim do minimalnej wartości. Obecnie coraz częściej wypierany przez bardziej złożone algorytmy.
- RMSProp – nowszy od SGD, obecnie częściej stosowany algorytm. w RMSProp nachylenie spadku gradientu jest wykorzystywane tylko do określenia kierunku spadku, a sama wartość jest dostosowana osobno dla każdej cechy [31].
- Adam – algorytm wykorzystujący podobnie do RMSProp zmienny współczynnik uczenia dla każdej cechy, jednak w inny sposób koryguje wartość korekty, uwzględniając gradienty z przeszłości. [31]

Algorytm	Koszt na zbiorze treningowym	Koszt na zbiorze testowym
SGD	0.871	0.879
RMSProp	0,833	0,815
Adam	0,788	0,780

Tabela 7.1 Porównanie wyników funkcji straty dla sieci wytrenowanych poszczególnymi algorytmami optymalizacji

Z komentarzem [BK(38)]: Hands-On Machine Learning with Scikit-Learn & TensorFlow STR 298



Rys. 7.1 Porównanie zmiany funkcji kosztu poszczególnych algorytmów w kolejnych epokach

Jak można zauważyć w tabeli 7.1. najskuteczniejszym okazał się algorytm Adam, a najsłabszym SGD. Ponieważ wartość błędu na zbiorze walidacyjnym jest zbliżona do wartości na zbiorze treningowym, można założyć, że w żadnym przypadku nie występuje zjawisko przeuczenia.

Na rysunku 7.2 przedstawiono zmienność błędu w kolejnych epokach dla badanych algorytmów. Jak można zauważyć, algorytm Adam nie tylko okazał się najskuteczniejszy ale także najszybciej osiąga stosunkowo niskie wartości błędu, dlatego wybrano go do użytkowania w finalnym modelu.

#### 7.4. Wybór funkcji straty

We współczesnych modelach najczęściej wykorzystuje się błąd średniokwadratowy, czyli funkcję opisywaną wzorem:

$$X = \frac{1}{n} \sum_{i=1}^n (Y_i - \check{Y}_i)^2 \quad (7.1)$$

Lub funkcję średniego bezwzględnego błędu proporcjonalnego:

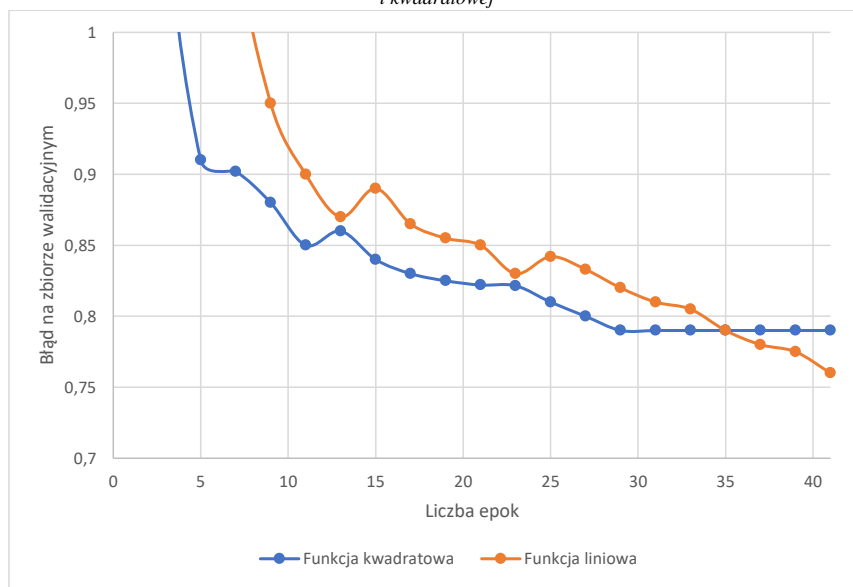
$$X = \frac{1}{n} \sum_{i=1}^n |Y_i - \check{Y}_i| \quad (7.2)$$

Gdzie:  $n$  – liczba próbek,  $Y_i$  – prawdziwy wynik,  $\check{Y}_i$  – wynik przewidziany przez model.

Poniżej zaprezentowano wyniki testów dla obu funkcji. Jako optymalizatora użyto wybranego wcześniej algorytmu Adam.

Funkcja	Koszt na zbiorze treningowym	Koszt na zbiorze walidacyjnym
Funkcja kwadratowa	0,788	0,780
Funkcja proporcjonalna	0,759	0,757

Tabela 7.2 Porównanie wyników funkcji straty dla sieci wytrenowanych przy użyciu funkcji proporcjonalnej i kwadratowej



Rys. 7.2 Porównanie zmiany straty sieci uczonych funkcjami proporcjonalną i kwadratową

Jak można zauważyć w tabeli 7.2., funkcja kwadratowa osiąga zauważalnie wyższe wartości. Danych z obu metod nie można jednak porównać bezpośrednio, ze względu na charakterystykę funkcji. W funkcji proporcjonalnej błąd 0,757 oznacza, że przeciętnie każda cecha wyliczona przez sieć, różni się o 0,757 od wartości prawdziwej. Ponieważ dane są ustandaryzowane jest to 0,757 odchylenia standardowego pierwotnej wartości. W funkcji kwadratowej błąd 0,780 oznacza różnicę pomiędzy produktem sieci a rzeczywistą wartością podniesioną do kwadratu. Zatem rzeczywista odległość to około 0,88 odchylenia standardowego – stanowczo więcej od funkcji proporcjonalnej.

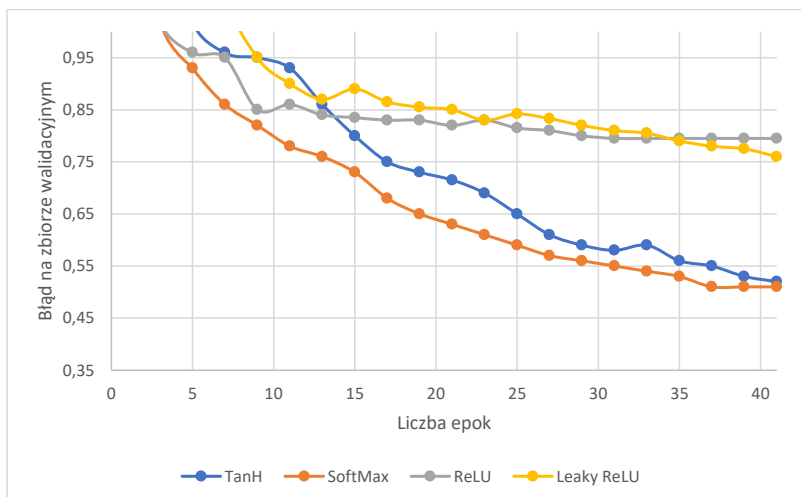
Na rysunku 7.3 przedstawiono zmienność błędu w kolejnych epokach dla obu funkcji. Jak widać, funkcja kwadratowa wyraźnie szybciej zbiega do niskich wartości, jednak na pewnym poziomie przestaje uczyć model. Funkcja proporcjonalna wydaje się nie duplikować tego problemu, przy dłuższym uczeniu prawdopodobnie można by osiągnąć jeszcze lepsze wyniki, dlatego wybrano funkcję błędu absolutnego do zastosowania w finalnym modelu.

## 7.5. Wybór funkcji aktywacji

We współczesnych sieciach, dla najlepszych efektów, zazwyczaj stosuje się funkcje nieliniowe. Poniżej zaprezentowano wyniki dla testów czterech obecnie najszerzej używanych metod: SoftMax, tangens hiperboliczny, ReLU i Leaky ReLU.

Funkcja	Koszt na zbiorze treningowym	Koszt na zbiorze walidacyjnym
SoftMax	0,511	0,508
Tangens hiperboliczny	0,521	0,522
ReLU	0,796	0,795
Leaky ReLU	0,766	0,761

*Tabela 7.3 Porównanie wyników funkcji straty dla sieci wytrenowanych przy użyciu poszczególnych funkcji aktywacji*



Rys. 7.3 Porównanie zmiany straty sieci uczonych przy użyciu poszczególnych funkcji aktywacji

Jak można zauważyć w tabeli 7.3, funkcje SoftMax i tangensa hiperbolicznego osiągając wyraźnie lepsze rezultaty.

Na rysunku 7.4. przedstawiono zmienność błędów w kolejnych epokach dla modeli uczonych poszczególnymi funkcjami. Jak można zauważyć, funkcja softmaxa i tangensa hiperbolicznego, sprawdzają się zdecydowanie lepiej niż ReLU i LeakyReLU. SoftMax osiąga nieznacznie lepszy wynik niż tangens hiperboliczny, jednak poprawa po każdej z epok spada wyraźnie szybciej w jej przypadku. Patrząc na tendencję spadku błędów, można podejrzewać, że przy dłuższym uczeniu funkcja SoftMax osiągnie gorsze rezultaty. Dlatego do implementacji w finalnym modelu wykorzystano funkcję tangensa hiperbolicznego.

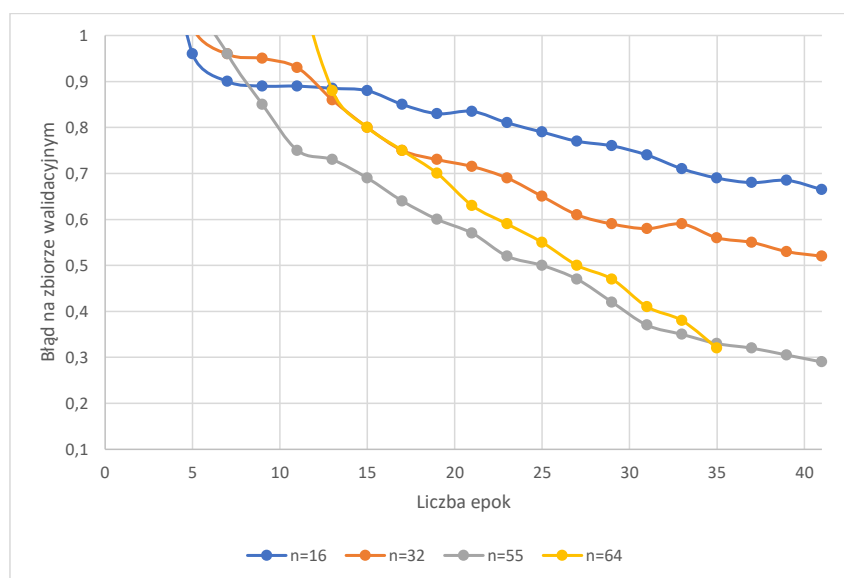
## 7.6. Wybór liczby neuronów w warstwach ukrytych

Zwiększenie liczby neuronów warstw może znacząco wpłynąć na pojemność informacyjną sieci, a co za tym idzie, zwiększyć dokładność predykcji. Jednak z drugiej strony, większy model zawiera więcej parametrów, a co za tym idzie wymaga większej mocy obliczeniowej i może znacząco spowolnić uczenie. w literaturze, zazwyczaj rozważa się wartości od 10 do 256 neuronów w warstwach ukrytych.

Ponieważ badany model zawiera 2 warstwy LSTM, możliwe jest użycie różnych licznosci jednostek w każdej warstwie. w podobnych przypadkach opisywanych w literaturze, stosowane są zazwyczaj równe licznosci dla wszystkich warstw rekurencyjnych. Poniżej zaprezentowano wyniki testów dla sieci zawierających po 16, 32, 55 (liczba cech wejściowych) i 64 neuronów w każdej warstwie LSTM.

Liczba neuronów w warstwach ukrytych	Koszt na zbiorze treningowym	Koszt na zbiorze walidacyjnym
16	0,651	0,665
32	0,523	0,521
55	0,296	0,294
64	0,309	0,32

Tabela 7.4 Porównanie wyników funkcji straty dla wytrenowanych sieci zawierających od 16 do 64 jednostek w warstwach LSTM



Rys. 7.4 Porównanie zmiany straty dla sieci zawierających od 64 do 512 jednostek w warstwach LSTM

Jak można zauważyć w tabeli 7.4, im więcej jednostek, tym skuteczniejsze uczenie modelu. Zbliżone wyniki na zbiorach treningowych i walidacyjnych świadczą, że przeuczenie nie występuje nawet przy największych wartościach.

Na rysunku 7.5. przedstawiono zmienność błędów na zbiorze walidacyjnym w kolejnych epokach dla sieci z poszczególną liczbą neuronów. Jak można zauważyć, sieci z mniejszą liczbą jednostek szybciej osiągają niskie wartości kosztu, ale tylko do pewnego poziomu, aby osiągnąć niższy wynik potrzeba więcej neuronów. Przy największej badanej sieci  $n=64$ , proces uczenia nie został zakończony na 40 epoki, ponieważ czas uczenia przekroczył zakładane 3 godziny. Najliczniejszy model osiągnął minimalnie lepszy wynik niż model z  $n=55$  w tej samej epoce, jednak mając na uwadze znacznie dłuższy czas uczenia (ok 45% dłużej) zdecydowano się zastosować  $n=55$ .

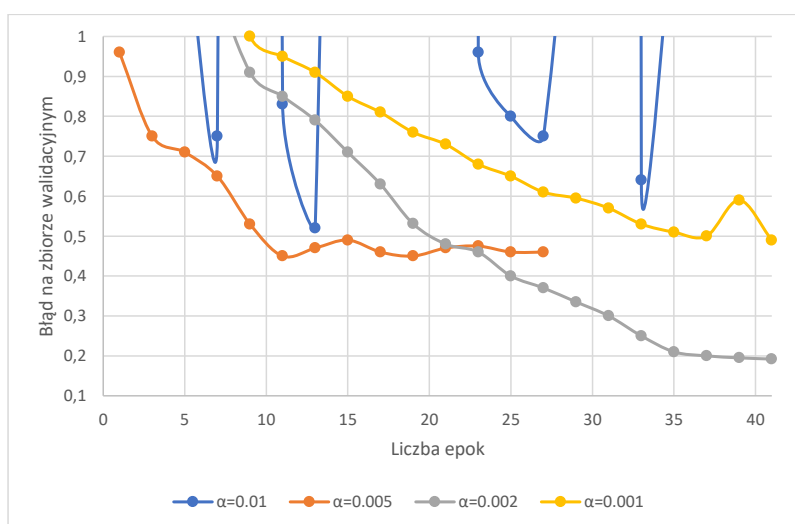


## 7.7. Wybór wartości początkowej współczynnika uczenia

Współczynnik uczenia jest parametrem liczbowym, odpowiadającym za wielkość korekty wag w każdej iteracji uczenia modelu. Zazwyczaj stosuje się stały współczynnik lub stopniowo zmniejsza go gdy sieć przestaje zmniejszać swój błąd. Poniżej zaprezentowano wyniki testów dla współczynnika uczenia ( $\alpha$ ) równego 0.01, 0.005, 0.002 oraz 0.001.

Wartość współczynnika uczenia	Koszt na zbiorze treningowym	Koszt na zbiorze walidacyjnym
0.01	1,35	1,37
0.005	0,46	0,457
0.002	0,192	0,197
0.001	0,496	0,501

Tabela 7.5 Porównanie wyników funkcji straty dla wytrenowanych sieci przy różnych wartościach współczynnika uczenia



Rys. 7.5 Porównanie zmiany straty dla wytrenowanych sieci przy różnych wartościach współczynnika uczenia

Jak można zauważyć w tabeli 7.5, najlepszy wynik został osiągnięty przy  $\alpha=0.002$ . Ponieważ eksperymenty zakończyły się na różnym etapie, trudno z tej informacji wyciągnąć odpowiednie wnioski.

Na rysunku 7.6 przedstawiono zmienność kosztu w kolejnych epokach dla różnych wartości współczynnika uczenia. Jak można zauważyć, dla wartości  $\alpha=0.01$ , sieć praktycznie się nie uczy, błąd osiąga zupełnie nieprzewidywalne wartości. Dla wartości  $\alpha=0.005$ , z kolei błąd osiąga dobre wyniki względnie szybko, jednak nie może zejść poniżej pewnej granicy. Dla wartości  $\alpha=0.002$  i  $\alpha=0.001$  model nie osiągnął granicy w czasie uczenia, jednak można zakładać, że dla  $\alpha=0.001$ , sieć uczyłaby się dłużej, ale osiągnęłaby niższy wynik.

Mając na uwadze powyższe porównanie oraz fakt, że współczynnik uczenia nie musi być parametrem stałym, zastosowano współczynnik  $\alpha=0.002$  na początku uczenia, oraz zmniejszony o połowę po każdych 10 epokach bez progresu.

## 7.8. Podsumowanie

W powyższym rozdziale wykonano szereg testów mających na celu wyłonienie optymalnych parametrów sieci. Ze względu na bardzo długi czas uczenia, oraz ograniczone zasoby, dla mniej istotnych parametrów (na przykład współczynnika nachylenia funkcji aktywacji) zastosowano domyślne wartości parametrów. Po wykonaniu, wybrano następujące parametry:

- algorytm uczenia sieci – Adam,
- funkcja kosztu – proporcjonalna,
- funkcje aktywacji – tangens hiperboliczny,
- liczba neuronów w warstwach ukrytych – 55,
- współczynnik uczenia – 0.002, zmniejszany o połowę po 10 epokach bez progresu.

Na tym etapie w żadnym przypadku nie zaobserwowano efektu przeuczenia modelu, zatem zdecydowano się nie implementować żadnych technik mających na celu uniknięcie go.

## 8. Uczenie i testy modelu

Dobór odpowiednich metod testowania modelu sieci rekurencyjnej nie jest zadaniem trywialnym, ponieważ w odróżnieniu od zwykłego klasyfikatora dla którego istnieją formalne metody testowanie nie istnieje uniwersalny sposób przetestowania modelu i określenia jego jakości.

### 8.1. Uczenie modelu

Sieć uczono przy użyciu parametrów dobranych w poprzednim rozdziale, zastosowano jednak inne kryteria końca nauki. Model uczy się maksymalnie 250 epok, bez limitu czasowego, jednak uczenie kończy się, gdy przez kolejne 15 epok model nie poprawi błędu na zbiorze walidacyjnym o co najmniej 0.02.

### 8.2. Walidacja krzyżowa

Metoda walidacji krzyżowej jest jednym z najpopularniejszych sposobów testowania klasyfikatorów. Ponieważ przy predykcji stanu pogodowego dla następnej godziny model działa jak klasyfikator, można przeprowadzić ten test, a jego wynik miarodajnie pokazuje jak dokładnie działa prognozowanie o jedną godzinę w przód.

Najczęstszą metodą walidacji krzyżowej jest metoda k-krotna, w której trenujemy i testujemy model na k różnie podzielonych zbiorach, tak aby każdy element zbioru był tylko raz

w zbiorze walidacyjnym. Zazwyczaj stosuje się walidację 10-krotną, ale ze względu na czasochłonność procesu uczenia zastosowano walidację 4-krotną.

Iteracja	Błąd na zbiorze treningowym	Błąd na zbiorze walidacyjnym
1	0,117	0,108
2	0,109	0,105
3	0,105	0,104
4	0,115	0,109

*Tabela 8.1 Wyniki walidacji krzyżowej dla kolejnych iteracji*

Jak można zauważyć w tabeli 8.1, wynik jest zbliżony dla każdego podziału danych i wynosi średnio 0,1065 odchylenie standardowe. Należy mieć na uwadze, że jest to błąd przy wyliczeniu prognozy dla najbliższej godziny, z każdą kolejną powinien on rosnąć. Na podstawie walidacji krzyżowej nie można jednak stwierdzić do jakiego poziomu urośnie strata na przestrzeni 72 godzin.

Dane w tabeli 8.1 pokazują, że w każdym przypadku błąd na zbiorze walidacyjnym jest zbliżony, a zazwyczaj nawet niższy niż błąd na zbiorze treningowym. Wyszło z tej informacji wniosek, że w modelu nie wystąpił problem przeuczenia. w związku z tym w dalszych etapach porzucono ostatecznie możliwość stosowania technik zapobiegania temu zjawisku.

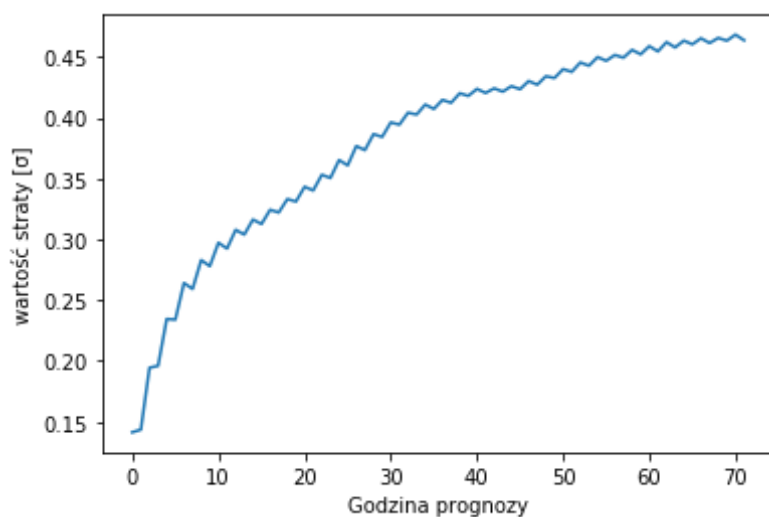
### 8.3. Testy dokładności modelu w kolejnych prognozowanych godzinach

W celu przetestowania skuteczności modelu w kolejnych godzinach zdecydowano się wyliczyć prognozę dla kolejnych 72 godzin i porównać ją z poprawnymi pomiarami.

Posłużono się użytym wcześniej algorytmem podziału na serie o wybranej długości. Dane testowe podzielono na serie długości 143 (aby wyliczyć pierwszy pomiar należy dysponować 71 poprzednimi). Aby przyspieszyć procedurę testową dane skompresowano analogicznie jak dane treningowe z założeniem  $n=64$ , w ten sposób liczba próbek zmniejszyła się 64 krotnie, jednocześnie zachowując  $64/143$  pojemności informacyjnej. w ten sposób otrzymano tensor o wymiarach [14475, 55, 143].

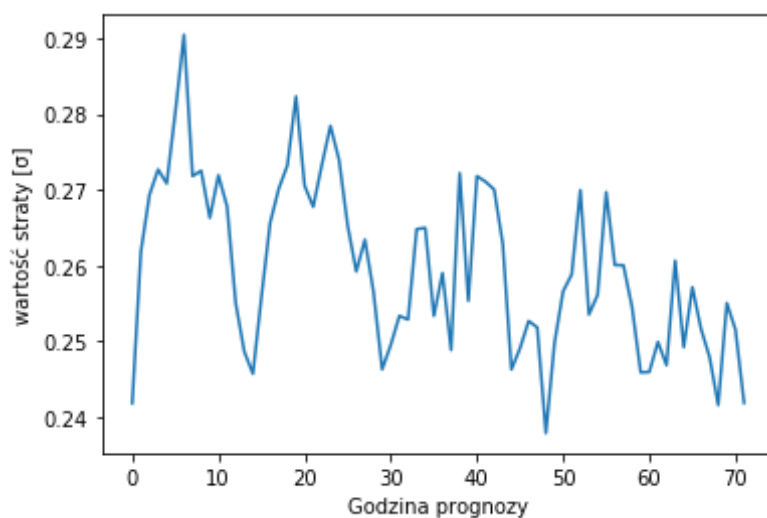
### 8.4. Badanie błędu całego wektora prognozy w czasie

Na wykresie 8.1 można zaobserwować że, zgodnie z przewidywaniami błąd rośnie wraz z czasem, w funkcji zbliżonej do logarytmicznej. Po 24 godzinach osiąga wartość 0,35 odchylenia standardowego, po 48 godzinach 0,41 odchylenia standardowego, a po 72 godzinach 0,45 odchylenia standardowego.



Rys. 8.1 Porównanie błędu modelu w kolejnych prognozowanych godzinach

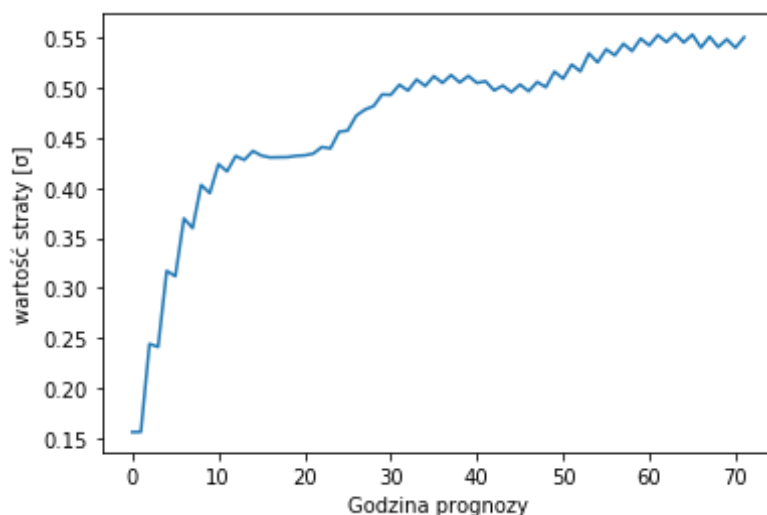
### 8.5. Badanie prognozy opadów w czasie



Rys. 8.2 Porównanie błędu prognozy opadów w kolejnych prognozowanych godzinach

Jak można zauważyć na wykresie 8.2 błąd w prognozie opadów rośnie tylko do około 6 godziny pomiaru gdzie osiąga 0,29 odchylenia standardowego błędu. w kolejnych godzinach wydaje się wahać w nierównych odstępach czasu i stopniowo maleć. w 72 godzinie błąd wynosi 0,24 odchylenia standardowego wynoszącego 0,501mm, czyli około 0,12mm.

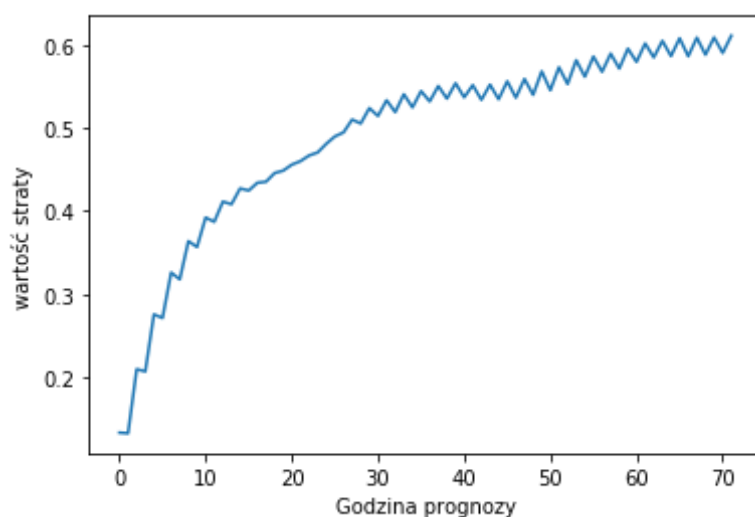
## 8.6. Badanie prognozy temperatury w czasie



Rys. 8.3 Porównanie błędu prognozy temperatury w kolejnych prognozowanych godzinach

Na wykresie 8.3 dostrzeżono, że błąd temperatury rośnie z czasem, analogicznie do całego wektora, jednak osiągając znacznie większe wartości. Odchylenie standardowe dla temperatury wynosi  $4,55\text{ }^{\circ}\text{C}$ . w 72 godzinie prognozy model osiąga błąd na poziomie 0,55 odchylenia standardowego czyli około  $2,5\text{ }^{\circ}\text{C}$ .

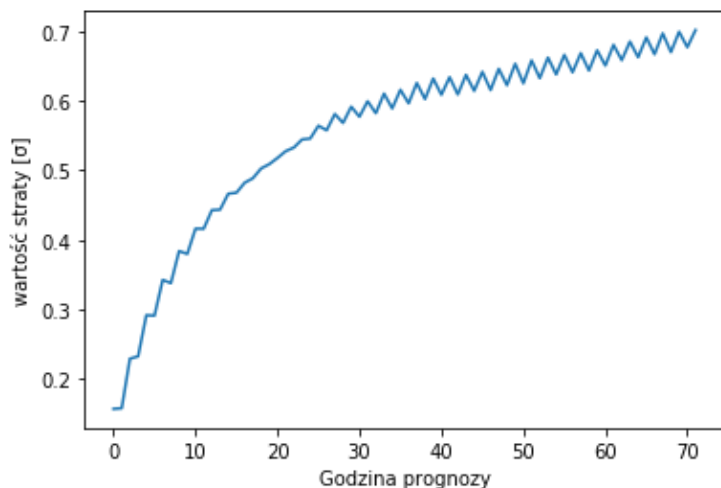
## 8.7. Badanie prognozy temperatury mokrego termometru w czasie



Rys. 8.4 Porównanie błędu prognozy temperatury mokrego termometru w kolejnych prognozowanych godzinach

Jak można zaobserwować na wykresie 8.4, skuteczność prognoz temperatury mokrego termometru jest bardzo zbliżona do prognozy temperatury, jednak osiąga nieznacznie wyższy błąd 0,61 odchylenia standardowego, które jest równe 4,17 °C, czyli około 2,54 °C.

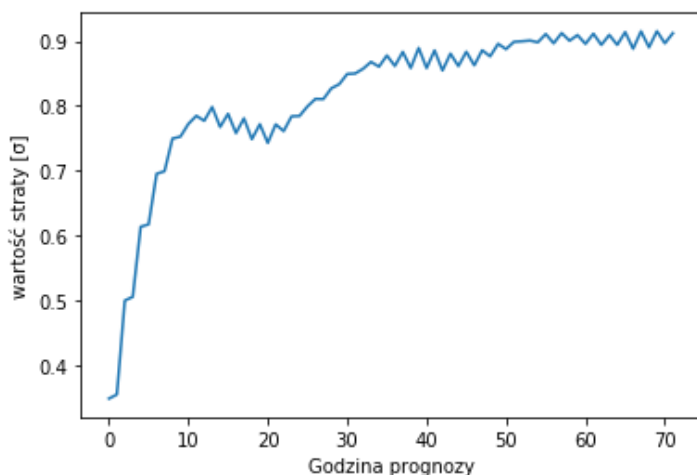
### 8.8. Badanie prognozy temperatury w punkcie rosy



Rys. 8.5 Porównanie błędu prognozy temperatury w punkcie rosy w kolejnych prognozowanych godzinach

Po przyjrzeniu się rysunkowi 8.5, można zaobserwować, że prognoza temperatury w punkcie rosy jest zbliżona do prognozy temperatury i temperatury mokrego termometru. Po 72 godzinach model osiąga 0,69 odchylenia standardowego błędu, który wynosi 4,42 °C, co odpowiada 3,04 °C błędowi.

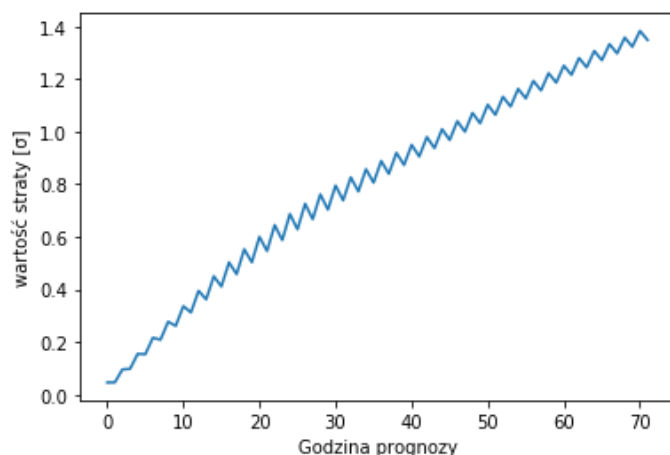
### 8.9. Badanie średniego błędu prognozy wilgotności względnej



Rys. 8.6 Porównanie błędu prognozy wilgotności w kolejnych prognozowanych godzinach

Jak można zauważyć na rysunku 8.6, błąd wilgotności rośnie z czasem, jednak zdaje się nieznacznie spadać po pełnych dobach prognoz. Może to świadczyć o pewnej prostej zależności dobowej, którą sieć jest w stanie odtworzyć, na przykład, że wilgotność każdego dnia o danej godzinie jest podobna. Błąd osiąga wartość 0,91 odchylenia standardowego, wynoszącego 11,58%, czyli około 10,53%

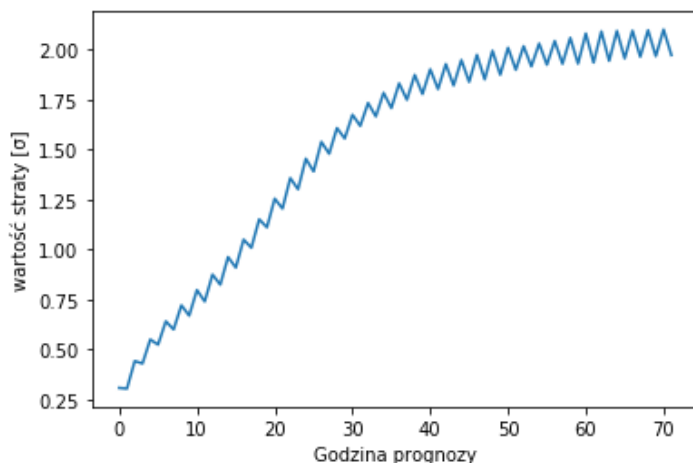
#### 8.10. Badanie średniego błędu prognozy ciśnienia atmosferycznego



Rys. 8.7 Porównanie błędów prognozy ciśnienia atmosferycznego w kolejnych prognozowanych godzinach

Z danych zamieszczonych na rysunku 8.7 można wyciągnąć wniosek, że błąd ciśnienia rośnie prawie liniowo, osiągając po 72 godzinach wartość 1,42 odchylenia standardowego wynoszącego 12,61 hPa. Oznacza to błąd na poziomie 17,9 hPa.

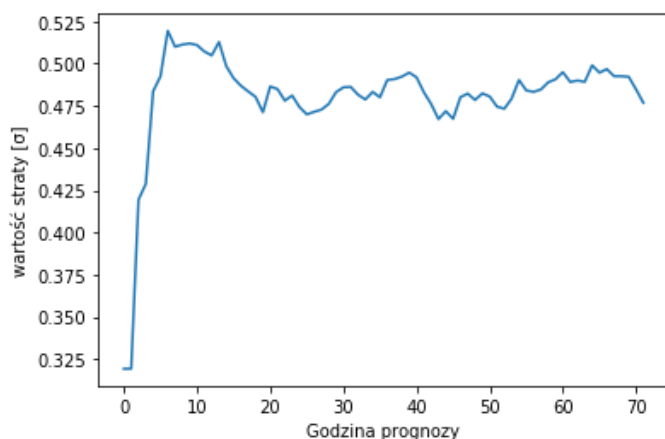
#### 8.11. Badanie średniego błędu prognozy prędkości wiatru



Rys. 8.8 Porównanie błędów prognozy prędkości wiatru w kolejnych prognozowanych godzinach

Na rysunku 8.8 dostrzeżono, że błąd rośnie w podobny sposób jak w przypadku pozostałych parametrów, osiągając w 72 godzinie wartość 2,08 odchylenia standardowego wynoszącego 6,4 węzła, czyli około 13,31 węzła błędu.

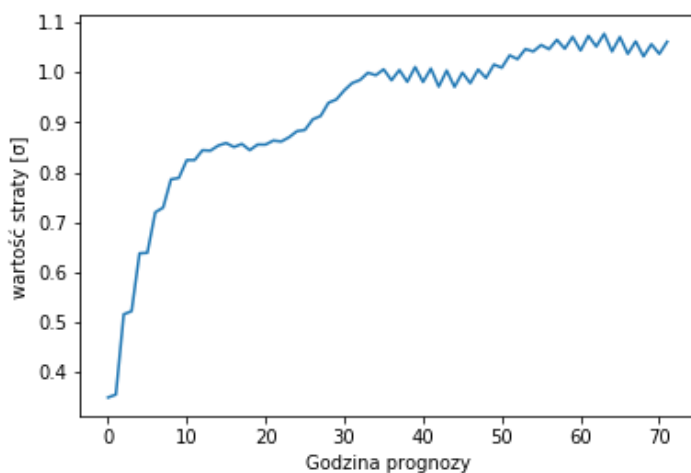
#### 8.12. Badanie średniego błędu prognozy nasłonecznienia



Rys. 8.9 Porównanie błędu prognozy nasłonecznienia w kolejnych prognozowanych godzinach

Jak można zaobserwować na wykresie 8.9, błąd modelu gwałtownie rośnie od drugiej godziny prognozy i osiąga wartość 0,513 odchylenia standardowego. w dalszych godzinach błąd nie rośnie już, ale utrzymuje zbliżoną wartość 0,480 odchylenia standardowego wynoszącego 0,32 godziny, co oznacza że model myli się średnio o 0,15 godziny nasłonecznienia.

#### 8.13. Badanie średniego błędu prognozy widoczności

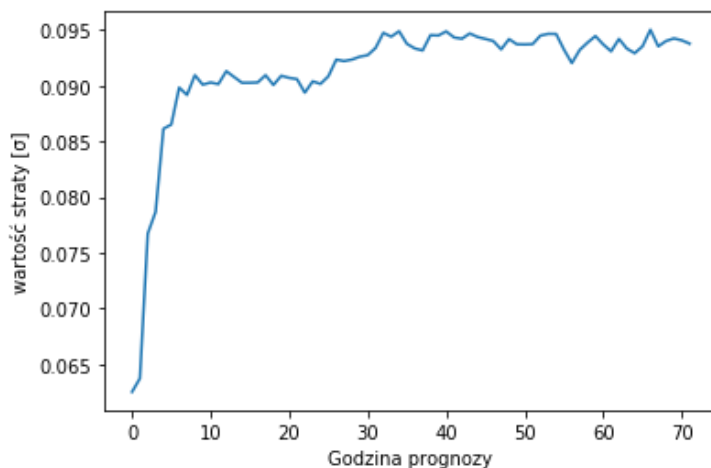


Rys. 8.10 Porównanie błędu prognozy widoczności w kolejnych prognozowanych godzinach



Z danych zamieszczonych na rysunku 8.10 można wysnuć wniosek, że błąd widoczności osiąga wartość 1,04 odchylenia standardowego w 72 godzinie pomiarów. Odchylenie standardowe odpowiada w tym przypadku 1577,96 metrom widoczności, co oznacza błąd na poziomie 1635,28 metra.

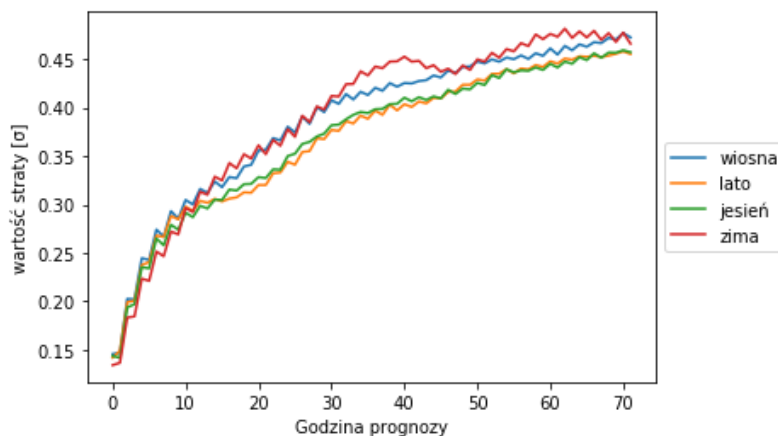
#### 8.14. Badanie średniego błędu kodu synoptycznego



Rys. 8.11 Porównanie błędu prognozy kodu synoptycznego w kolejnych prognozowanych godzinach

Jak można zaobserwować na rysunku 8.11, kod synoptyczny jest najlepiej prognozowanym parametrem. Błąd rośnie podobnie do pozostałych parametrów ale w szczytowym momencie osiąga tylko 0.093 – czyli ponad 5 krotnie mniej niż cały wektor. Trudno szukać przyczyny tego zjawiska ale możliwe, że po prostu kod jest bardzo stabilny lub łatwy w przewidzeniu.

#### 8.15. Porównanie skuteczności modelu w różnych sezonach roku.



Rys. 8.12 Porównanie błędu prognozy w zależności od sezonu

Na rysunku 8.12 widać, że skuteczność modelu jest zbliżona we wszystkich porach roku. Najlepsze wyniki daje prognoza na lato i jesień, a najgorsze na wiosnę i zimę. Interesujący jest wynik dokładności pomiarów w okresie zimowym, gdzie błąd okresowo maleje i rośnie. Prawdopodobnie wynika to ze specyfiki zimy, gdzie pogoda jest bardziej powtarzalna.

#### **8.16. Wnioski**

Powyżej przedstawione testy dają czytelny obraz skuteczności modelu w różnych przypadkach. Wyraźnie widać, że model nie prognozuje wystarczająco skutecznie większości parametrów. Najlepiej można zobaczyć to na przykładzie temperatury, gdzie błąd osiąga wartości 2,51 °C, przy średniej dobowej amplitudzie około 7 °C. Nawet proste modele synoptyczne potrafiłyby przewidzieć wynik z podobną dokładnością. Jednak z drugiej strony, część parametrów jest prognozowana z zadowalającą dokładnością, na przykład ilość opadów jest prognozowana z dobrą dokładnością nawet w ostatnich godzinach testu.

Dobrze sprawdza się jednak prognozowanie kodu synoptycznego, gdzie błąd nawet po 3 dniach przekracza zaledwie 15%.

Dla większości elementów stanu pogodowego, błąd rośnie logarytmicznie lub w funkcji zbliżonej do logarytmicznej. Jest to zgodne z przewidywaniami. w związku z charakterystyką funkcji logarytmicznej, można wysnuć hipotezę, że wytrenowany model mógłby lepiej sprawdzić się dla dłuższych prognoz – dla późniejszych godzin błąd powinien rosnać coraz wolniej.

Dla wszystkich parametrów z wyjątkiem wilgotności nie zaobserwowano cyklicznych zmian błędów co 24 godziny. Taka zależność mogłaby świadczyć, że sieć ma problem z odwzorowaniem zależności pomiędzy porami dnia, jednak w tym wypadku model nauczył się rozwiązywać ten problem, a błąd wynika z innych, mniej trywialnych zależności.

Przy pomiarach części parametrów, można zauważyć pewne anomalie, polegające na wzroście błędów tylko co drugą godzinę. Jest to bardzo dobrze widoczne na wykresach, gdzie można zaobserwować wzrost błędów krokowy jako charakterystyczne „zębki”. Najbardziej widoczne jest to od 1 do 15 godziny pierwszej i drugiej doby prognozy. Przyczyna tego zjawiska jest nieznana, jednak ze względu na to, że anomalie nie są znaczące, można uznać je za charakterystykę działania modelu, a nie znaczący problem.

Biorąc po uwagę powyższe wnioski, można wysnuć tezę, że zaimplementowany model nie nadaje się do prognoz 3 dniowych. Dokładność dla krótszych prognoz jest jednak wyraźnie lepsza, dla większości parametrów prognoza 24 godzinna jest około dwukrotnie skuteczniejsza niż 72 godzinna.

## 9. Podsumowanie

W ramach niniejszej pracy, został stworzony model głębokiego uczenia prognozujący pogodę krótkotrwałą dla obszaru Irlandii. Model został wytrenowany na odpowiednio dobranym i przygotowanym zbiorze danych, następnie dokonano doboru optymalnych hiperparametrów i wytrenowano model. Nauczona sieć została dokładnie przetestowana pod kątem skuteczności prognozy każdego z elementów stanu pogodowego.

Obecne rezultaty modelu nie są zadowalające, błąd na poziomie 0,46 odchylenia standardowego po 72 godzinach prognozy jest znacznie wyższy od innych współczesnych modeli. Ze względu na charakterystykę wzrostu błędu, model niewiele lepiej sprawdzi się także dla krótszych prognoz.

Rozważając model jako prognozujący na krótszy okres czasu, to dla przykładu, średni błąd wynoszący 0,35 odchylenia standardowego po 24 godzinach jest wyraźnie mniejszy, jednak w tak krótkim okresie czasu predykcja jest znacznie prostsza więc od poprawnie działającego modelu można się spodziewać znacznie niższych wartości błędu. Wnioskując na podstawie tych danych można stwierdzić, że w żadnym oknie czasowym mniejszym niż 72 godziny, model nie sprawdzi się w prognozie całego wektora stanu pogody.

Pomimo słabej dokładności całego przewidywanego wektora, zaskakująco dobrze działa predykcja kodu synoptycznego – który opisuje słownie stan pogody. Działanie systemu można więc ograniczyć do skutecznego przewidywania typu pogody, jaki wystąpi w pożądaną chwilę.

Analogicznie do kodu synoptycznego, niskie wartości błędu uzyskiwane są także dla ilości opadów – w tym wypadku jednak, nie zaobserwowano wyraźnego wzrostu błędu w czasie. Możliwe, że jest to związane z kodem synoptycznym, który definiuje czy opady w ogóle występują. Mając te informacje na uwadze, można założyć, że model mógłby być skutecznie wykorzystywany do predykcji występowania i nasilenia opadów atmosferycznych.

Projekt systemu można ocenić jako częściowo udany, ponieważ na bieżącym etapie zadowalająca jest jedynie predykcja kodu synoptycznego i opadów. Nie udało się zrealizować pierwotnego celu zakładającego skuteczną prognozę całego wektora pogodowego, nie mniej jednak obecne wyniki pozwalają na użycie modelu w szczególnych zastosowaniach – prognozie opadów i typu pogody na najbliższe 72 godziny.

### 9.1. Możliwe przyczyny niskiej skuteczności modelu

#### 9.1.1. Trudna do przewidzenia charakterystyka klimatu Irlandii

W pracach naukowych na których wzorowano się podczas realizacji tego projektu podobne modele osiągały o wiele lepsze wyniki nawet przy dużo mniejszych zbiorach treningowych i prostszych architekturach sieci. Jednak wszystkie rozważane prace dotyczyły innych lokalizacji, nie znaleziono żadnego rozwiązania działającego na obszarze Irlandii ani

pobliskich terenów. Można z tego wyciągnąć wniosek, że sam klimat wybranego obszaru jest na tyle nieprzewidywalny, że należałoby podejść do niego w inny sposób. w literaturze [5] jest on opisywany jako stabilny i mało zmienny, możliwe że dlatego wszystkie odchylenia są znacznie trudniejsze do przewidzenia niż w innych miejscach.

### **9.1.2. Niewłaściwy dobór parametrów i architektury**

Dobór hiperparametrów został przeprowadzony tak, aby jak zoptymalizować błąd modelu w pierwszej godzinie prognozy. Możliwe że przy innym zestawie parametrów, na przykład znacznie większym rozmiarze warstwy ukrytej błąd rósłby wolniej w kolejnych godzinach prognozy lub stabilizował się na niższej wartości. Dobór parametrów sieci nie jest jednak zadaniem trywialnym i każde podejście ma swoje zalety jak i wady.

W literaturze opisanych jest wiele modeli prognozujących pogodę, z czego można znaleźć co najmniej kilka różnych architektur modeli głębokiego uczenia. Skuteczność wybranego rozwiązania jest bardzo mocno powiązana z wybranym zbiorem danych, dlatego możliwe jest, że dla zadanych danych sieć z dwoma warstwami LSTM sprawdza się gorzej niż inna, na przykład sieć konwolucyjna.

## **9.2. Dalsze możliwości rozwoju projektu**

Projekt można uznać za skończony, jednak istnieje szereg możliwości dalszego rozwoju, które można zastosować w przyszłości.

### **9.2.1. Porównanie skuteczności modelu z innymi architekturami sieci**

Tak jak opisano w poprzednim podrozdziale, istnieje co najmniej kilka innych podejść do głębokiego uczenia w tym temacie. Jako dalszy rozwój projektu można by zaimplementować część z nich i porównać otrzymane rezultaty.

### **9.2.2. Znalezienie lepszej metody doboru parametrów sieci**

Na etapie hiperparametrów wartości były dobierane w celu optymalizacji prognozy na kolejną godzinę. Dobrym rozwiązaniem byłoby znalezienie sposobu na optymalizację błędu na cały okres prognozy, ponieważ jak można zauważyć, błąd w pierwszej godzinie faktycznie jest niewielki, ale drastycznie rośnie w kolejnych.

### **9.2.3. Użycie modelu na innym zbiorze danych w innej lokalizacji**

Większość analizowanych prac naukowych skupia się na terenie Indii lub Chin. Po znalezieniu odpowiedniego zbioru danych treningowych, istniałaby możliwość porównania wprost jakości rozwiązania w stosunku do innych. Możliwe jest również, że w innych lokalizacjach model radziłby sobie wyraźnie lepiej, przez inne własności klimatu.

## Bibliografia

- [1] Alley, R. B., Emanuel, K. A., Zhang, F. (2019). Advances in weather prediction. *Science*, 363(6425): 342–344.
- [2] Stern, H., & Davidson, N. E. (2015). Trends in the skill of weather prediction at lead times of 1-14 days. *Quarterly Journal of the Royal Meteorological Society*, 141(692): 2726–2736.
- [3] „Top 500 supercomputers list” [Online]:  
<http://www.top500.org> [Data uzyskania dostępu: 8.05.2020]
- [4] „Strona światowej organizacji meteorologicznej” [Online]:  
<https://www.wmo.int/pages/prog/www/DPS/GDPS-Supplement5-AppI-4.html>  
[Data uzyskania dostępu: 23.04.2020]
- [5] Allaby, M. *ENCYCLOPEDIA OF weather and climate Revised Edition Volume I*, New York Facts on File 2002, ISBN-13: 978-0816063505.
- [6] „Awiacja” Instytutu Meteorologii i Gospodarki Wodnej [Online]:  
<http://awiacja.imgw.pl/index.php> [Data uzyskania dostępu 23.04.2020]
- [7] Oficjalna strona „VisualCrossing” [Online]:  
<https://www.visualcrossing.com/resources/blog/how-do-we-create-our-multi-location-weather-forecast/> [Data uzyskania dostępu 25.04.2020]
- [8] Smith D.L., Zuckerberg F.L., Schaefer J.T., Rasch G.E. (1986) Forecast Problems: The Meteorological and Operational Factors. In: Ray P.S. (eds) *Mesoscale Meteorology and Forecasting*. American Meteorological Society, Boston, MA
- [9] „Encyklopedia Britannica -History Of Weather Forecasting ” [Online]:  
<https://www.britannica.com/science/weather-forecasting/History-of-weather-forecasting>  
[Data uzyskania dostępu 25.04.2020]
- [10] Krasnopolsky, V. M., & Fox-Rabinovitz, M. S. (2006). Complex hybrid models combining deterministic and machine learning components for numerical climate modeling and weather prediction. *Neural Networks*, 19(2): 122–134.
- [11] Oficjalna strona National Oceanic and Atmospheric Administration [Online]:  
<https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forecast-system-gfs>  
[Data uzyskania dostępu 29.04.2020]

- [12] Oficjalna strona Europejskiego Centrum Prognoz Średnioterminowych [Online]: <https://www.ecmwf.int/en/forecasts/documentation-and-support> [Data uzyskania dostępu 29.04.2020]
- [13] Zangl G. ICON: The Icosahedral Nonhydrostatic Modelling Framework of DWD and MPI-M
- [14] Oficjalna strona Deutscher Wetterdienst [Online]: [https://www.dwd.de/EN/research/weatherforecasting/num\\_modelling/01\\_num\\_weather\\_prediction\\_modells/icon\\_description.html](https://www.dwd.de/EN/research/weatherforecasting/num_modelling/01_num_weather_prediction_modells/icon_description.html) [Data uzyskania dostępu 29.04.2020]
- [15] Oficjalna strona Met Office [Online]: <https://www.metoffice.gov.uk/research/approach/modelling-systems/unified-model/weather-forecasting> [Data uzyskania dostępu 29.04.2020]
- [16] „Electric Utility gets ahead of the weather with new forecasting models” [Online]: [https://earthnetworks.com/Portals/0/pdf/IBM\\_Deep\\_Thunder\\_Case\\_Study.pdf](https://earthnetworks.com/Portals/0/pdf/IBM_Deep_Thunder_Case_Study.pdf) [Data uzyskania dostępu 29.04.2020]
- [17] Oficjalna strona „IBM DeepThunder” [Online]: <https://www.ibm.com/ibm/history/ibm100/us/en/icons/deepthunder/> [Data uzyskania dostępu 29.04.2020]
- [18] Oficjalna strona Z danymi Instytutu Meteorologii i Gospodarki Wodnej [Online]: <https://danepubliczne.imgw.pl> [Data uzyskania dostępu 05.05.2020]
- [19] Oficjalna strona Instytutu „Met eireann” [Online]: <https://www.met.ie/climate/available-data/historical-data> [Data uzyskania dostępu 05.05.2020]
- [20] „kaggle.com” [Online]: <https://www.kaggle.com/conorrot/irish-weather-hourly-data> [Data uzyskania dostępu 05.05.2020]
- [21] Abhishek, K., Singh, M. P., Ghosh, S., & Anand, A. (2012). Weather Forecasting Model using Artificial Neural Network. *Procedia Technology*, 4: 311–318.
- [22] Maqsood, I., Khan, M., & Abraham, A. (2004). An ensemble of neural networks for weather forecasting. *Neural Computing and Applications*, 13(2).
- [23] Greff, K., Srivastava, R. K., Koutnik, J., Steunebrink, B. R., & Schmidhuber, J. (2017). LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, 28(10): 2222–2232.

- [24] Salman, A. G., Kanigoro, B., & Heryadi, Y. (2015). Weather forecasting using deep learning techniques. *2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*.
- [25] Liu T., Fang S., Zhao Y., Wang P., Zhang J. Implementation of Training Convolutional Neural Networks
- [26] Zhao, B., Li, X., Lu, X., & Wang, Z. (2018). A CNN-RNN Architecture for Multi-Label Weather Recognition. *Neurocomputing*.
- [27] Shi X., Chen Z., Wang H., Yeung D. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting
- [28] Le X., Ho H., Lee G. (2019). Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting. *Water*, 11(7): 1387.
- [29] Zaytar, Akram M., Amrani E., Chaker. (2016). Sequence to Sequence Weather Forecasting with Long Short-Term Memory Recurrent Neural Networks. *International Journal of Computer Applications*. 143: 7-11.
- [30] Salman, A. G., Heryadi, Y., Abdurahman, E., & Suparta, W. (2018). Single Layer & Multi-layer Long Short-Term Memory (LSTM) Model with Intermediate Variables for Weather Forecasting. *Procedia Computer Science*: 135, 89–98.
- [31] Géron A. Hands-On Machine Learning with Scikit-Learn & TensorFlow, Helion, 2019, ISBN: 978-1492032595.
- [32] „ruder.io” [Online]: <https://runder.io/optimizing-gradient-descent/index.html#whichoptimizertouse> [Data uzyskania dostępu 20.05.2020]

## Spis rysunków

<i>Rys. 1.1 Zmiana skuteczności prognozy pogody 5 dniowej w kolejnych latach na przykładzie systemu GFS, gdzie: NH – półkula północna, SH – półkula południowa [2] .....</i>	<i>5</i>
<i>Rys. 3.1 Przykładowa prognoza pogody na 1 dzień z serwisu openweathermap.org .....</i>	<i>7</i>
<i>Rys. 3.2 Porównanie skuteczności prognozy modelu „Climate Forecast System” na przestrzeni czasu dla prognozy 1 dniowej, 2-4 dniowej, 5-7 dniowej, 8-10 dniowej i 11-14 dniowej[2] .....</i>	<i>9</i>
<i>Rys. 4.1 Przykładowy wynik działania modelu GFS .....</i>	<i>11</i>
<i>Rys. 5.1 Stacje pomiarowe z największą liczbą pomiarów .....</i>	<i>15</i>
<i>Rys. 5.2 Zmiany średnich temperatur w wybranych hrabstwach na przestrzeni roku .....</i>	<i>16</i>
<i>Rys. 5.3 Średnia dobowa ilość opadów w poszczególnych stacjach pomiarowych .....</i>	<i>17</i>
<i>Rys. 5.4 Rozkład temperatur na przestrzeni lat .....</i>	<i>19</i>
<i>Rys. 5.5 Porównanie zmian średnich temperatur na przestrzeni całego roku dla lat 1989 i 2017 .....</i>	<i>20</i>
<i>Rys. 5.6 Średnia dobowa ilość opadów w poszczególnych latach .....</i>	<i>21</i>
<i>Rys. 5.7 Odchylenie standardowe od średnich dobowych opadów w poszczególnych latach .....</i>	<i>22</i>
<i>Rys. 5.8 Uśredniona widoczność dla kwietnia, maja i czerwca w latach od 2008 do 2013 ....</i>	<i>23</i>
<i>Rys. 6.1 Schemat komórki LSTM .....</i>	<i>26</i>
<i>Rys. 6.2 Algorytm przetwarzania danych treningowych .....</i>	<i>28</i>
<i>Rys. 6.3 Uproszczony schemat algorytmu konwersji danych na serie .....</i>	<i>31</i>
<i>Rys. 6.4 Schemat budowy sieci .....</i>	<i>33</i>
<i>Rys. 7.2 Porównanie zmiany funkcji kosztu poszczególnych algorytmów w kolejnych epokach .....</i>	<i>36</i>
<i>Rys. 7.3 Porównanie zmiany straty sieci uczonych funkcjami proporcjonalną i kwadratową .....</i>	<i>37</i>
<i>Rys. 7.4 Porównanie zmiany straty sieci uczonych przy użyciu poszczególnych funkcji aktywacji .....</i>	<i>39</i>
<i>Rys. 7.5 Porównanie zmiany straty dla sieci zawierających od 64 do 512 jednostek w warstwach LSTM .....</i>	<i>40</i>
<i>Rys. 7.6 Porównanie zmiany straty dla wytrenowanych sieci przy różnych wartościach współczynnika uczenia .....</i>	<i>41</i>
<i>Rys. 8.1 Porównanie błędu modelu w kolejnych prognozowanych godzinach .....</i>	<i>44</i>
<i>Rys. 8.2 Porównanie błędu prognozy opadów w kolejnych prognozowanych godzinach .....</i>	<i>44</i>
<i>Rys. 8.3 Porównanie błędu prognozy temperatury w kolejnych prognozowanych godzinach .....</i>	<i>45</i>
<i>Rys. 8.4 Porównanie błędu prognozy temperatury mokrego termometru w kolejnych prognozowanych godzinach .....</i>	<i>45</i>
<i>Rys. 8.5 Porównanie błędu prognozy temperatury w punkcie rosy w kolejnych prognozowanych godzinach .....</i>	<i>46</i>
<i>Rys. 8.6 Porównanie błędu prognozy wilgotności w kolejnych prognozowanych godzinach .....</i>	<i>46</i>



<i>Rys. 8.7 Porównanie błędu prognozy ciśnienia atmosferycznego w kolejnych prognozowanych godzinach .....</i>	47
<i>Rys. 8.8 Porównanie błędu prognozy prędkości wiatru w kolejnych prognozowanych godzinach .....</i>	47
<i>Rys. 8.9 Porównanie błędu prognozy nasłonecznienia w kolejnych prognozowanych godzinach .....</i>	48
<i>Rys. 8.10 Porównanie błędu prognozy widoczności w kolejnych prognozowanych godzinach .....</i>	48
<i>Rys. 8.11 Porównanie błędu prognozy kodu synoptycznego w kolejnych prognozowanych godzinach .....</i>	49
<i>Rys. 8.12 Porównanie błędu prognozy w zależności od sezonu .....</i>	49

## **Spis tabel**

<i>Tabela 7.1 Porównanie wyników funkcji straty dla sieci wytrenowanych poszczególnymi algorytmami optymalizacji .....</i>	35
<i>Tabela 7.2 Porównanie wyników funkcji straty dla sieci wytrenowanych przy użyciu funkcji proporcjonalnej i kwadratowej .....</i>	37
<i>Tabela 7.3 Porównanie wyników funkcji straty dla sieci wytrenowanych przy użyciu poszczególnych funkcji aktywacji .....</i>	38
<i>Tabela 7.4 Porównanie wyników funkcji straty dla wytrenowanych sieci zawierających od 16 do 64 jednostek w warstwach LSTM .....</i>	40
<i>Tabela 7.5 Porównanie wyników funkcji straty dla wytrenowanych sieci przy różnych wartościach współczynnika uczenia .....</i>	41
<i>Tabela 8.1 Wyniki walidacji krzyżowej dla kolejnych iteracji .....</i>	43