

Predykcja aktywności fizycznej z wykorzystaniem uczenia maszynowego

Zuzanna Gez

Cel eksperymentu



Celem projektu jest opracowanie modelu predykcyjnego, który będzie w stanie dokładnie prognozować dzienną liczbę kroków użytkownika. Kluczowym aspektem jest identyfikacja najbardziej efektywnej metody opartej na analizie szeregów czasowych.

Wybór optymalnego rozwiązania zostanie dokonany na podstawie dokładności i niezawodności prognoz, które zostaną zweryfikowane za pomocą wskaźników jakości, takich jak RMSE, MAE oraz R2.

Dane projektowe

Do analizy wykorzystano dane dotyczące dziennej liczby kroków, pozyskane z aplikacji Health, zbierającej informacje o aktywności fizycznej z urządzeń IoT. Dane pochodzą z okresu od 9 grudnia 2017 roku do 31 grudnia 2024 roku. W celu wzbogacenia analizy, dane te zostały połączone z historycznymi informacjami meteorologicznymi, pobranymi z API weather.visualcrossing.com.

Zgromadzony zbiór danych zawiera daty pomiarów, poziom aktywności fizycznej, warunki pogodowe oraz dodatkowe zmienne kalendarzowe.

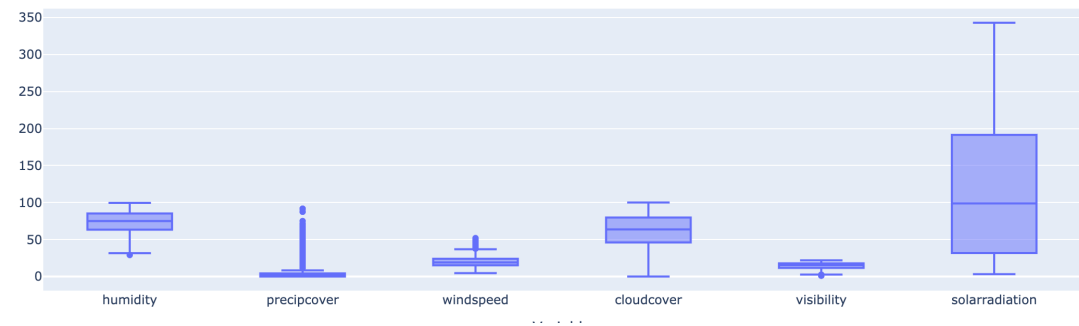
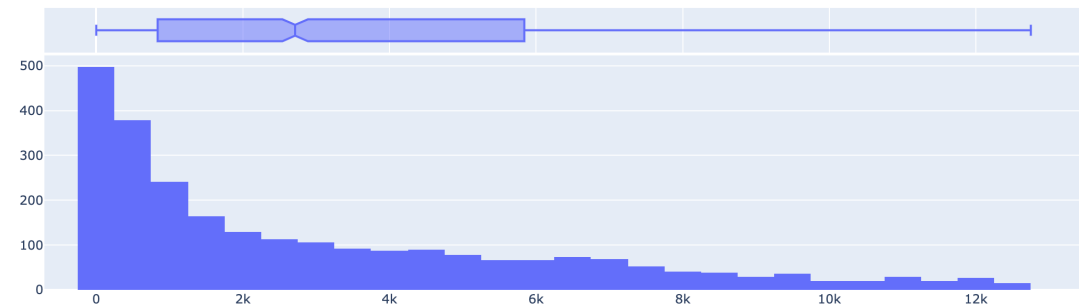
	date	value	month	year	day_of_week	is_wknd	temp	humidity	precipcover	windspeed	cloudcover	visibility	solarradiation
0	2017-12-09	617.0	12	2017	6	1	0.7	81.8	0.00	14.5	71.9	12.3	28.6
1	2017-12-10	1325.0	12	2017	7	1	0.9	78.0	0.00	20.3	77.7	14.1	25.4
2	2017-12-11	2559.0	12	2017	1	0	3.5	77.2	8.33	21.1	81.6	17.0	16.9
3	2017-12-12	7467.0	12	2017	2	0	7.3	69.6	0.00	26.5	78.7	17.7	21.1
4	2017-12-13	4866.0	12	2017	3	0	1.5	73.0	0.00	20.4	59.3	15.4	24.4

Wstępna ocena i przygotowanie danych

Dane dotyczące aktywności fizycznej charakteryzowały się losowymi brakami, silną asymetrią oraz obecnością wartości odstających. Zastosowanie interpolacji pozwoliło na uzyskanie spójnego zbioru danych z ograniczoną liczbą ekstremalnych wartości.

Z kolei dane meteorologiczne cechowały się niskim rozproszeniem. Jednak w przypadku promieniowania słonecznego i opadów zaobserwowano silną asymetrię oraz liczne wartości odstające.

Analiza nie wykazała istotnego wpływu pogody na liczbę kroków. Zmienne pogodowe i inne cechy niezwiązane bezpośrednio z historią aktywności zostały odrzucone.



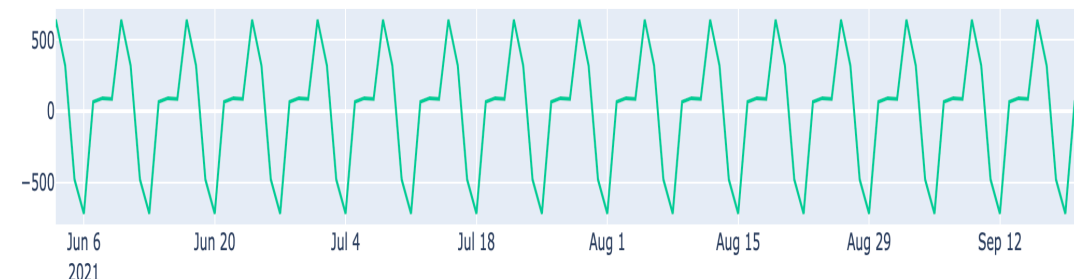
date	1	-0.200571	0.1142662	0.989788	-672.313μ	-389.0867μ	0.05777737	0.03633984	-0.04592717	-0.005406168	0.09410429	0.1245555	-0.002286855
value	-0.200571	1	0.01131422	-0.2031529	-0.07274563	-0.07967451	-0.01298953	0.003199871	-0.098523	-0.03653601	-0.03798196	-0.06899806	-0.006103778
month	0.1142662	0.01131422	1	-0.02801867	0.003905553	0.003862646	0.1660446	0.2209907	-0.04717927	-0.1411128	0.0389064	-0.09482137	-0.1127404
year	0.989788	-0.2031529	-0.02801867	1	-0.001242641	-934.7615μ	0.03448651	0.004272616	-0.03989188	0.01461617	0.08874966	0.1392774	0.01430775
day_of_week	-672.313μ	-0.07274563	0.003905553	-0.001242641	1	0.8661933	0.002525689	-0.002365582	-0.01094499	0.01363077	0.02915344	-0.003377504	0.005956904
is_wknd	-389.0867μ	-0.07967451	0.003862646	-934.7615μ	0.8661933	1	0.001651268	-0.003814629	-0.009122291	0.02128812	0.02793092	0.003725771	0.004857115
temp	0.05777737	-0.01298953	0.1660446	0.03448651	0.002525689	0.001651268	1	-0.4809816	-0.1302828	-0.07272067	-0.4227801	0.5634335	0.7412652
humidity	0.03633984	0.003199871	0.2209907	0.004272616	-0.002365582	-0.003814629	-0.4809816	1	0.3593145	-0.08022436	0.6578767	-0.7189742	-0.7943984
precipcover	-0.04592717	-0.098523	-0.04717927	-0.03989188	-0.01094499	-0.009122291	-0.1302828	0.3593145	1	0.1971522	0.3283717	-0.1827061	-0.2632076
windspeed	-0.005406168	-0.03653601	-0.1411128	0.01461617	0.01363077	0.02128812	-0.07272067	-0.08022436	0.1971522	1	0.1046545	0.2358141	-0.05793871
cloudcover	0.09410429	-0.03798196	0.0389064	0.08874966	0.02915344	0.02793092	-0.4227801	0.6578767	0.3283717	0.1046545	1	-0.4460584	-0.6696874
visibility	0.1245555	-0.06899806	-0.09482137	0.1392774	-0.003377504	0.003725771	0.5634335	-0.7189742	-0.1827061	0.2358141	-0.4460584	1	0.631436
solarradiation	-0.002286855	-0.006103778	-0.1127404	0.01430775	0.005956904	0.004857115	0.7412652	-0.7943984	-0.2632076	-0.05793871	-0.6696874	0.631436	1
date	value	month	year	day_of_week	is_wknd	temp	humidity	precipcover	windspeed	cloudcover	visibility	solarradiation	

Wstępna ocena i przygotowanie danych

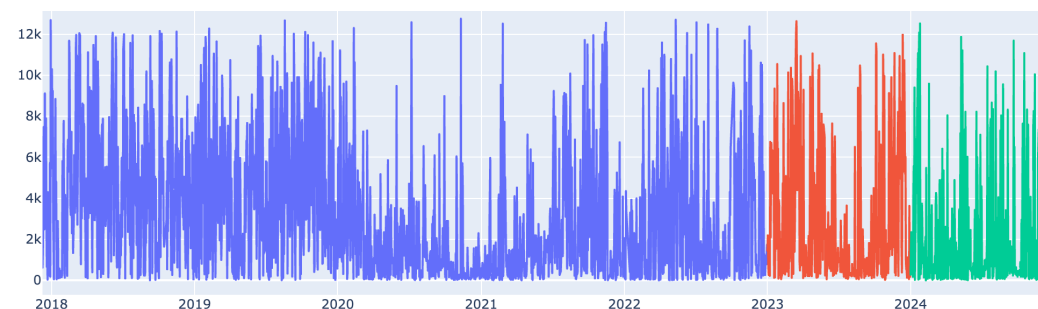
Dane ujawniły wyraźne tygodniowe wzorce sezonowe.

Kluczowe okazały się cechy oparte na wcześniejszych wartościach liczby kroków. Wykorzystano opóźnienia, różnice w czasie oraz średnie kroczące, które wygładziły krótkoterminowe wahania i ułatwiły identyfikację długoterminowych trendów.

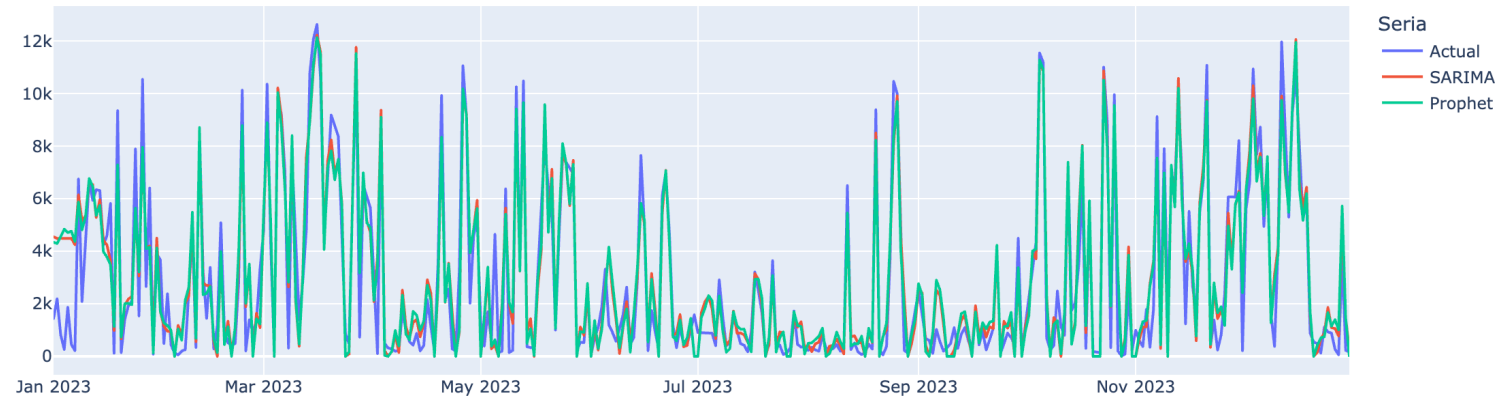
Zbiory danych podzielono chronologicznie na trzy części: treningowy (2017–2022), walidacyjny (2023) oraz testowy (2024).



value	1	0.3589983	0.2417101	0.2837195	0.4061302	0.4023511	0.3952675	0.6394377	0.6363597	0.6025146
value_lag_1	0.3589983	1	0.3592289	0.2364602	0.506988	0.4718283	0.4359948	0.1134578	0.09172282	0.1351399
value_lag_2	0.2417101	0.3592289	1	0.2310426	0.5214748	0.4799234	0.4431263	-0.03210648	0.03837698	0.08288107
value_lag_7	0.2837195	0.2364602	0.2310426	1	0.5563798	0.5165356	0.463891	0.03386404	0.05455843	0.04541888
value_rolling_mean_15	0.4061302	0.506988	0.5214748	0.5563798	1	0.9514499	0.8601084	-0.1172487	-0.0791268	0.07255873
value_rolling_mean_22	0.4023511	0.4718283	0.4799234	0.5165356	0.9514499	1	0.9196029	-0.09462119	-0.09611838	0.05662339
value_rolling_mean_43	0.3952675	0.4359948	0.4431263	0.463891	0.8601084	0.9196029	1	-0.05576664	-0.06182656	-0.06052033
value_diff_9	0.6394377	0.1134578	-0.03210648	0.03386404	-0.1172487	-0.09462119	-0.05576664	1	0.5301263	0.4505436
value_diff_15	0.6363597	0.09172282	0.03837698	0.05455843	-0.0791268	-0.09611838	-0.06182656	0.5301263	1	0.4586124
value_diff_35	0.6025146	0.1351399	0.08288107	0.04541888	0.07255873	0.05662339	-0.06052033	0.4505436	0.4586124	1



Modelowanie



Do prognozowania dziennej liczby kroków wykorzystano modele Prophet oraz SARIMA.

Skuteczność oceniono za pomocą miar błędów: RMSE, MAE i R^2 , a eksperymenty przeprowadzono na zbiorze walidacyjnym. Oba modele osiągnęły porównywalne wyniki, jednak Prophet wykazał nieznacznie niższe wartości RMSE i MAE, a także krótszy czas trenowania.

	Model	RMSE	MAE	R^2	Training Time (s)
0	SARIMA	1139.015000	831.631901	0.880097	9.336507
1	Prophet	1145.250441	847.222418	0.878781	0.203931

Ewaluacja

W trakcie walidacji krzyżowej SARIMA uzyskała lepsze wyniki, co sugeruje jej większą stabilność.

	Model	RMSE	MAE	R ²
0	SARIMA	1077.888518	837.473097	0.863140
1	Prophet	1085.709269	858.383655	0.859287

Na danych testowych model SARIMA dobrze odwzorowuje ogólny trend, jednak spadek wartości R² sugeruje nieznacznie gorszą skuteczność modelu w porównaniu do wyników uzyskanych na zbiorze treningowym.

	Dataset	RMSE	MAE	R ²
0	Validation	1125.87	818.77	0.88
1	Test	1169.35	861.16	0.84

Wnioski



Analiza wpływu pogody, w tym zmiennych takich jak temperatura, wilgotność, opady i prędkość wiatru, nie wykazała istotnego wpływu na liczbę kroków użytkowników. Okazuje się, że aktywność fizyczna jest bardziej zależna od indywidualnych nawyków i rutyn, niż od warunków atmosferycznych.

Modele szeregów czasowych, oparte na wcześniejszych wartościach aktywności, wykazały dużą skuteczność w przewidywaniu liczby kroków. Wykorzystanie lagowych wartości oraz różnic w czasie pozwoliło uchwycić sezonowe wzorce w aktywności.

Model SARIMA okazał się bardziej precyzyjny w prognozowaniu liczby kroków, szczególnie w odniesieniu do wychwytywania sezonowych fluktuacji oraz zmienności aktywności. Wykazał się większą stabilnością i skutecznością w porównaniu do modelu Prophet.