
HACKATON DATA FOR CITY

Zespół 100 Twarzy Grzybiarzy

Damian Bagiński, Jędrzej Sokołowski, Malwina Wojewoda

Spis treści

Rozwiązanie	2
Jakie problemy rozwiązujemy?	2
Co zrobiliśmy?	3
Jak to zrobiliśmy?.....	6
Jak można rozwinąć nasz pomysł?	8
Jak można wdrożyć nasze rozwiązanie?	9

Rozwiązanie

Zaproponowany przez nas pomysł przesłaliśmy jako folder zip oraz znajduje się w repozytorium: https://github.com/malwina0/Hackaton_DataForCity.

Jakie problemy rozwiązujemy?

Naszą motywacją było umożliwienie warszawiakom łatwiejszego wyszukiwania interesujących ich wydarzeń. Jasne jest, że zazwyczaj mieszkańcy bardziej zainteresowani są co dzieje się w ich okolicy. Właśnie dlatego wizualizacja imprez na mapie oraz możliwość wybrania, w jakim promieniu od wskazanego punktu chcemy je zobaczyć zdecydowanie ułatwi znalezienie czegoś ciekawego dla danego odbiorcy. A im łatwiej znajdzie coś takiego, tym większa szansa, że zdecyduje się uczestniczyć, co jest oczywistą korzyścią dla organizatorów. Poza tym, promowanie lokalnych wydarzeń pozytywnie wpływa na sąsiedzką integrację oraz aktywizację mieszkańców Ursynowa, jak i pozostałych dzielnic. Co więcej, w [Załączniku nr 1 - Liście wyzwań miejskich](#) na stronie Hackatonu punkt 9. zawiera informację, że istnieje potrzeba stworzenia listy wszystkich wydarzeń z miasta z podziałem na dzielnice, co oznacza, że jest to ważna kwestia nie tylko dla nas, ale i dla Organizatorów.

Podczas inauguracji Hackatonu mieliśmy okazję przekonać się, że ważną kwestią dla mieszkańców jest dbanie o środowisko, ponieważ wiele z pytań młodych osób dotyczyło właśnie tego. Temat ten jest bliski również nam i chcielibyśmy zrobić coś w kierunku poprawy jakości powietrza. Edukowanie i uświadamianie prowadzą do zmian, dlatego stworzyliśmy „prognozę smogową”, która może wzbudzić dyskusję społeczną. Zanieczyszczone powietrze ma negatywny wpływ na zdrowie, dlatego taka prognoza może pomóc mieszkańcom ułożyć sobie plan dnia tak, żeby możliwie zredukować ekspozycję na smog i tym samym zadbać o siebie. Jest to szczególnie istotne w przypadku dzieci i osób cierpiących na choroby przewlekłe.

Ilustrując to na przykładzie: Mieszkanka Ursynowa ma małe dziecko i nie posiada samochodu. Chce wyjść z synkiem w miejsce w okolicy, ponieważ unika zatłoczonych miejsc, w tym komunikacji miejskiej. Zaznacza w naszej aplikacji swoje miejsce zamieszkania i wybiera promień 2 kilometrów, bo tyle będzie w stanie się przejść. Pokazuje jej się kilka wydarzeń, z czego najbardziej zaciekał ją plenerowy teatr lalek w parku, który odbywa się za 2 dni. Klika na nie na liście pod mapą, aby dowiedzieć się więcej. Ważne jest dla niej, żeby synek nie był narażony na zanieczyszczone powietrze, więc w zakładce Smog sprawdza prognozę smogu na kolejne dni. Okazuje się, że nie będzie przekraczał norm, dlatego może już wcześniej zaplanować sobie kolejne dni tak, by móc uczestniczyć. W trakcie przedstawienia

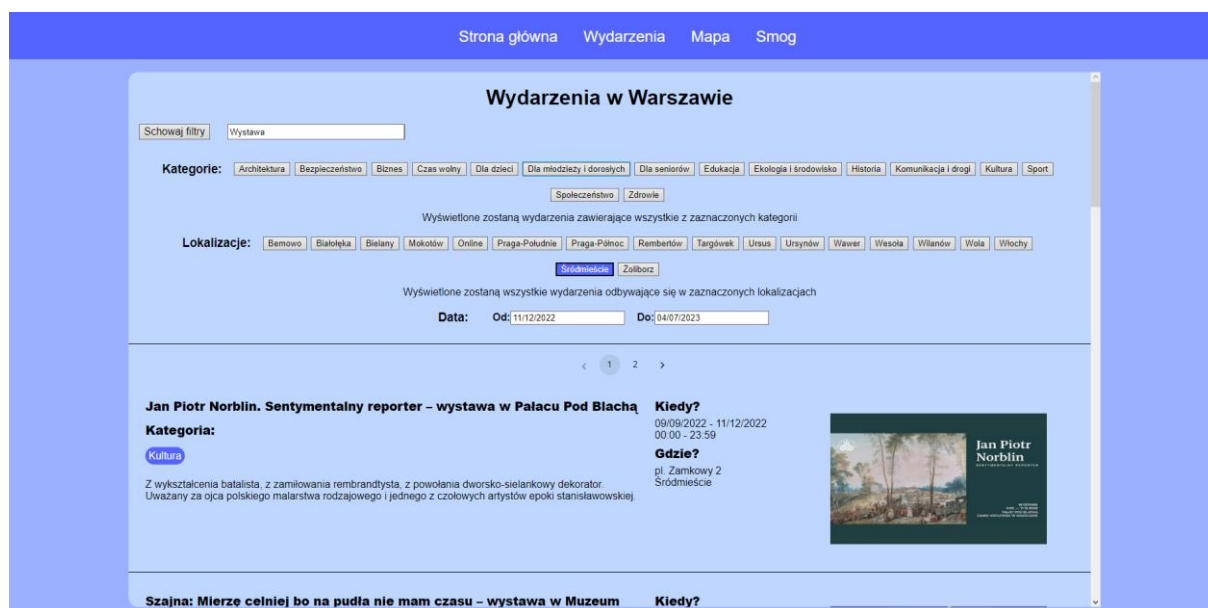
synek poznaje nowych kolegów, a jego mama sąsiadki i planują razem pójść na kolejne wydarzenia.

Co zrobiliśmy?

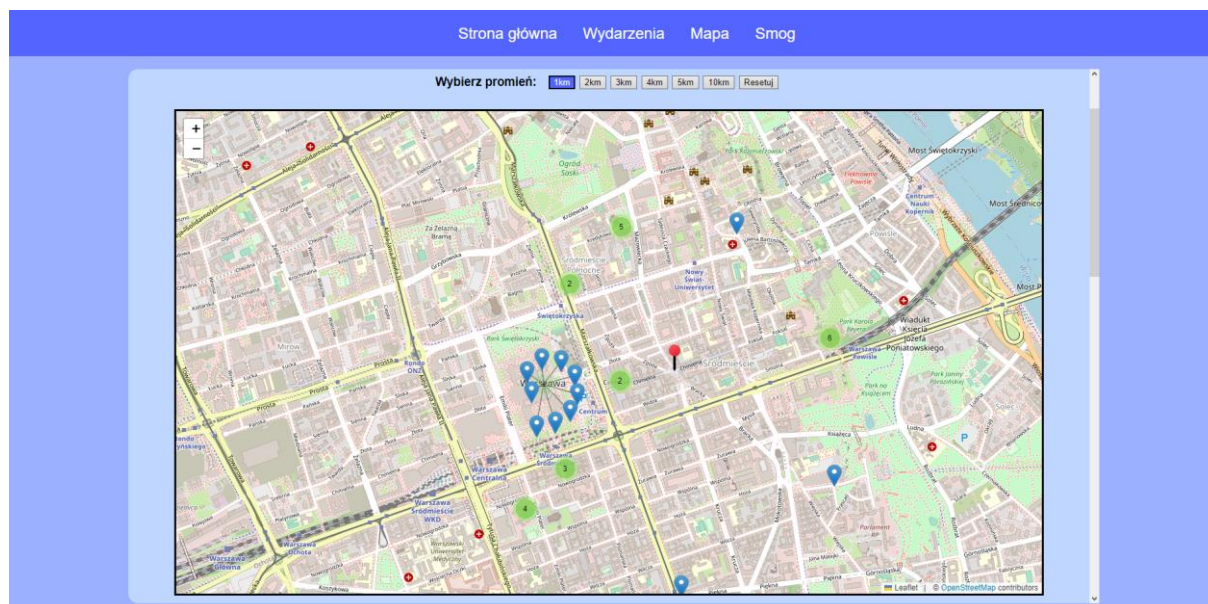
Naszą pracą konkursową jest aplikacja webowa, gdzie na stronie głównej wyjaśniamy krótko jej funkcjonalności.



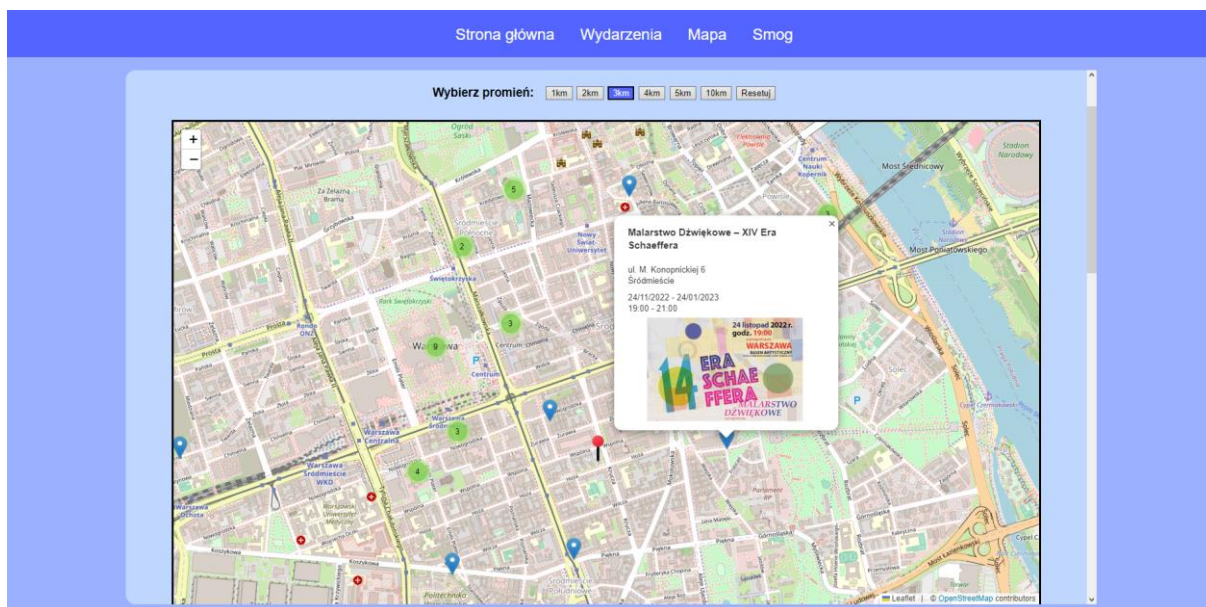
W pierwszej zakładce – Wydarzenia można zobaczyć listę najbliższych wydarzeń w Warszawie, domyślnie posortowanych rosnąco względem dat. Można filtrować je zgodnie z własnymi potrzebami: wybierając dzielnicę, kategorię oraz zakres dat, a także wyszukiwać po słowach kluczowych. Na stronie mieści się 10 wydarzeń, jeśli jest ich więcej należy przejść na kolejną stronę. Po kliknięciu na jedno z nich nastąpi przekierowanie na stronę internetową, na której przedstawione są jego szczegóły.



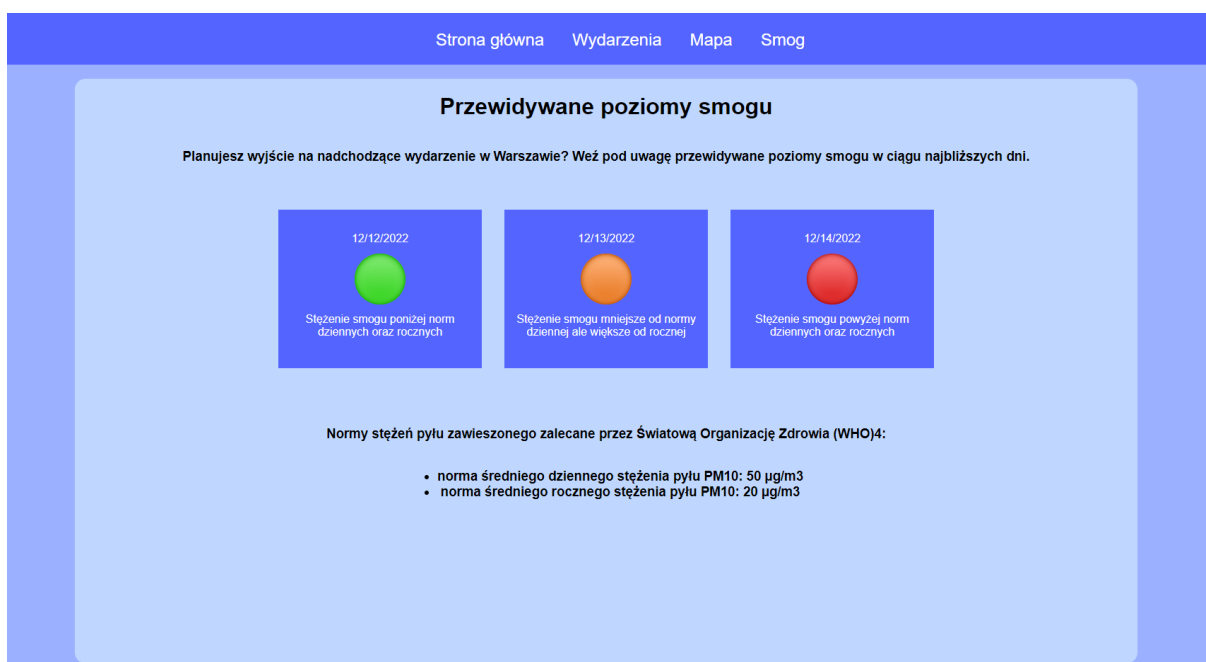
W osobnej zakładce znajduje się mapa, na której wyświetlone są zgrupowane wszystkie wydarzenia odbywające się w najbliższym czasie. Przybliżając i oddalając mapę możemy bliżej zobaczyć, gdzie dokładnie się znajdują. Ponadto istnieje możliwość wybrania dowolnego punktu na mapie oraz wartości promienia, w jakiej odległości od tego punktu mają nam się wyświetlić imprezy. Poniżej możemy także zobaczyć te filtrowane wydarzenia w postaci listy, analogicznej jak w zakładce Wydarzenia.



Po kliknięciu na jedno z wydarzeń wyświetli się o nim więcej informacji:



W zakładce Smog, stworzyliśmy prognozę zanieczyszczenia powietrza na kolejny dzień. Przewidywane są konkretne wartości pyłów zawieszonych, których średnica jest nie większa niż 1, 2.5 oraz 10 mikrometra, a także poziom dwutlenku azotu. Te zanieczyszczenia są bezpośrednio związane ze smogiem. W aplikacji zdecydowaliśmy się zilustrować jedynie czy dane te spełniają normy ustalone przez WHO, ponieważ uznaliśmy, że dla zwykłego odbiorcy to będzie czytelniejsza informacja niż konkretne liczby, z których samemu trzeba coś wywnioskować.



Jak to zrobiliśmy?

Strona została stworzona w Pythonie. Do części backendowej aplikacji został użyty webowy framework Django. Za część graficzną (frontendową) odpowiada javascriptowy framework React. Dodatkowo użyty został również język programowania CSS.

Dane dotyczące wydarzeń stworzona przez nas aplikacja pobiera bezpośrednio z [Otwarte dane - czyli dane po warszawsku](#).

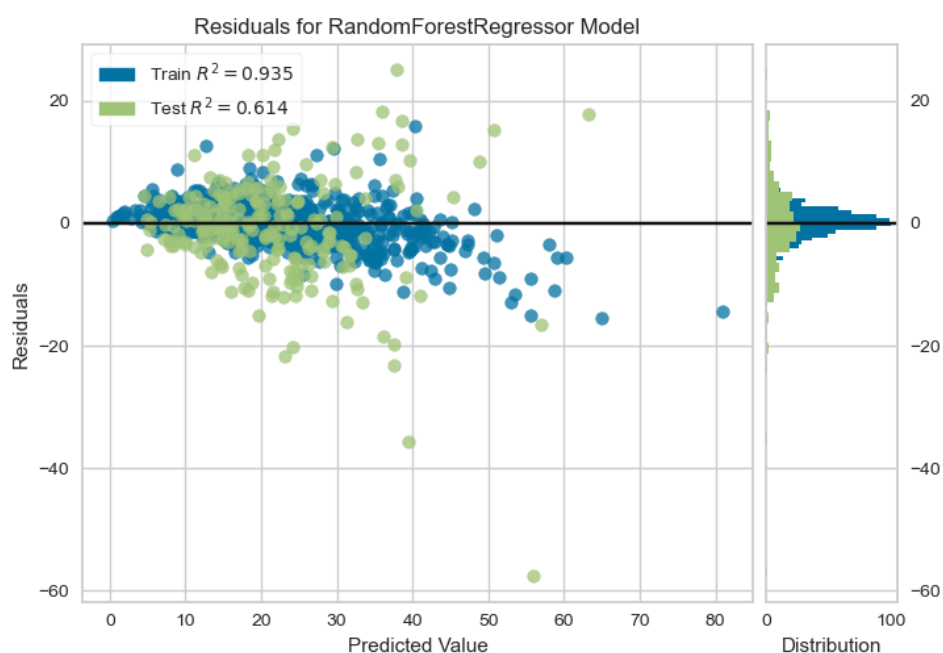
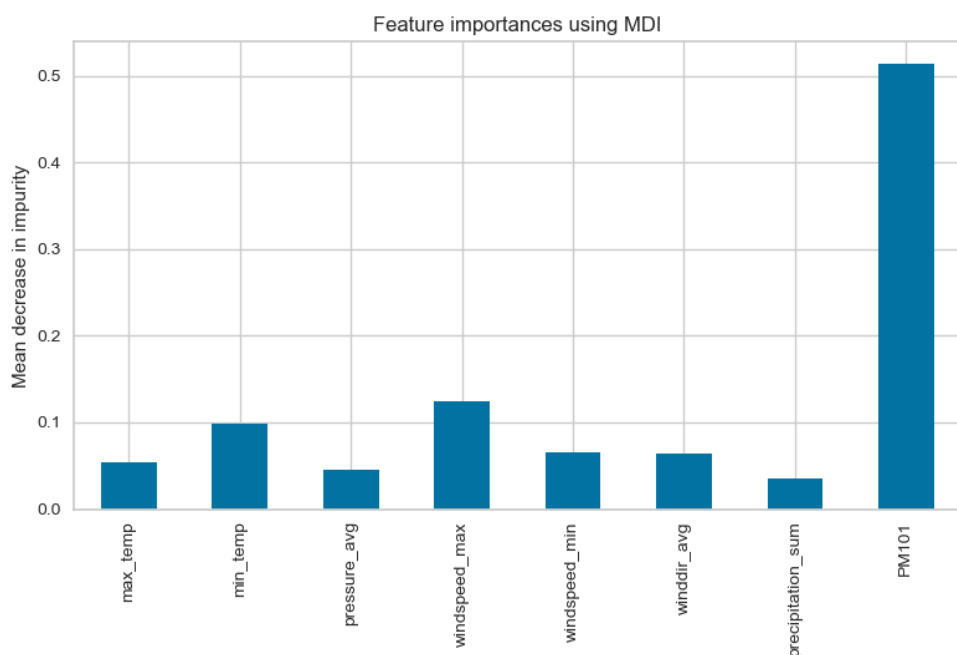
Model przewidujący zanieczyszczenie powietrza

Model 1:

Do wykonania modelu użyliśmy danych o zawartości pyłu zawieszonego PM10 ze strony [Głównego Inspektoratu Ochrony Środowiska](#). Dane te pochodzą ze stacji przy ulicy Wokalne. Zdecydowaliśmy się na tę, ponieważ zlokalizowana jest w środkowej części Ursynowa. Pobraliśmy dane dla okresu 3.12.2019 - 3.12.2022 r.

Dane pogodowe dla tego samego okresu pobraliśmy ze strony [Open-Meteo](#) i obejmują konkretnie datę, temperaturę, ciśnienie atmosferyczne, opady atmosferyczne, szybkość oraz kierunek wiatru.

Na podstawie tych danych stworzyliśmy skrypt wykonujący dwa modele regresyjne – Random Forest oraz xgBoost. Za każdym razem model wybiera najlepszy model na podstawie wartości R^2 . Parametr “verbose” przy funkcji run_model_training ustawiony na “True” pozwala na zobaczenie dystrybucji błędów modelu oraz błędów średniokwadratowych, dzięki czemu w przyszłości, jeżeli byłby (ten model) dalej rozwijany, łatwo można by było porównywać inne modele z różnymi hiperparametrami np. przy pomocy rozwiązania mlflow. Poniżej zamieszczamy 2 grafiki dotyczące modelu - ważność zmiennych (bardzo mocny udział opóźnionej wartości zmiennej objaśnianej) oraz rozkład reszt z modelu dla jednego z Foldów modelu (do prześledzenia wszystkich wystarczy zmienić parametr verbose na True w funkcji run_model_training.py).



Model 2:

Ze względu na granularność danych oraz trudności w prawidłowym przewidywaniu stężeń zanieczyszczeń na dłuższą metę, proponujemy podejście dwumodelowe. Pierwszy z nich to opisany wcześniej odpowiedzialny za przewidywanie skali smogu średnio dla całego dnia, a drugi umożliwiłby dokładniejszą ocenę sytuacji poprzez przewidywanie wartości związków szkodliwych do max 8 godzin wprzód.

Do stworzenia tego modelu wykorzystaliśmy dane z czujników stanu powietrza udostępnione na stronie [Otwarte dane - czyli dane po warszawsku](#). Strona udostępnia API

umożliwiające pobranie aktualnych danych z wielu czujników zlokalizowanych w różnych miejscach Warszawy. Zdecydowaliśmy się użyć tych ze stacji zlokalizowanej przy ulicy Warchałowskiego 8, ponieważ jest to czujnik znajdujący się w środkowej części Ursynowa. Oczywiście model można analogicznie zastosować do pozostałych sensorów, jednak przygotowaliśmy prototyp tylko na jednym z nich, z Ursynowa - dzielnicy, której dedykowany jest ten Hackaton.

Nie udało nam się zdobyć danych z tych czujników z poprzednich kilku dni, dlatego stworzyliśmy mechanizm, który pobiera i zapisuje dane z API regularnie co 10 minut (ponieważ co tyle czasu występuje aktualizacja). Być może w urzędzie już istnieje baza danych z tymi danymi, gdyby jednak nie to oszacowaliśmy, że w ciągu roku dane te zajmowałyby około 7,5 GB przy założeniu, że zostałyby one zapisywane w formacie txt. Kompresja pliku do zip pozwoliłaby na zmniejszenie tej liczby nawet o 90%.

Niestety ze względu na niewielką liczbę danych, nie byliśmy w stanie zbudować drugiego modelu, który byłby godny zaufania. Należałoby obserwować i pobierać dane jeszcze przez jakiś czas, by móc opracować model, który rzetelnie przewidywałby wartości zanieczyszczeń. Zamieściliśmy próby naszego rozwiązania w pliku [model_per_hour.py](#), znajdujący się w folderze modelowanie_na_godzine. Jako zbiór testowy służyły nam dane z 2.12.2022 r. O ile wyniki naszych prób dla PM10, PM2.5, PM1 nie wydają się być złe, o tyle przypadek z NO2 jest dużo trudniejszy. Potencjalnym rozwiązaniem tego problemu mogłoby być wytrenowanie zupełnie innych modeli dla poszczególnych zanieczyszczeń.

```
Mean averages for predicted values: PM10: 30.46169444444447 PM2.5 22.185611111111104 PM1 18.991375 NO2 24.57558333333333
Mean averages for real values: PM10: 26.31069444444445 PM2.5 21.70986111111111 PM1 14.46097222222222 NO2 15.99208333333333
```

Oprócz tego wykorzystaliśmy dane pogodowe, pochodzące z API udostępnianego przez stronę [Instytutu Meteorologii i Gospodarki Wodnej](#) i analogicznie jak w przypadku danych smogowych zbieraliśmy aktualne dane.

Okres zbierania danych to 26.11.2022 r. godzina 17:000 – 02.12.2022 r., godzina 11:50.

Dane połączone ze sobą znajdują się w pliku [data_merged.csv](#).

Jak można rozwinąć nasz pomysł?

O ile etap pozyskiwania aktualnych danych poprzez API został już przez nas zautomatyzowany, tak do pełnej automatyzacji brakuje, aby strona każdego dnia sama wykonywała modelowanie przy użyciu nowo pobranych danych.

Należałoby także zdobyć API predykujące pogodę na następne kilka dni, mogłoby to być używane już przez nas udostępniane przez stronę [Open-Meteo](#). Dzięki temu, można by

było przewidywać wartości smogu na wiele więcej dni wprzód. Oczywistym jest fakt, że im więcej wartości smogu chcemy przewidywać (np. na 2 tygodnie wprzód), tym ich wiarygodność się zmniejsza, jednak uważamy, że dalej byłaby to wartościowa informacja dla użytkownika.

Z powodu ograniczonego czasu proponujemy tu jedynie prototyp rozwiązania. Model drugi, przewidujący smog na kilka godzin wprzód, można ulepszyć poprzez korzystanie z danych z kilku dni wcześniej. Prognoza, którą zrobiliśmy jest uśrednioną dla całego dnia, lecz w przypadku, gdybyśmy posiadali więcej danych szczegółowych, co 10 minut, pewnie udałooby się stworzyć dobre przewidywania na poszczególne wartości nawet co godzinę.

Przewidywania zrobiliśmy jedynie dla jednej lokalizacji – czujnika przy ul. Warchałowskiego 8, jednak analogiczny model można wykonać dla wszystkich pozostałych, a także te znajdujące się niedaleko w jakiś sposób ze sobą powiązać.

Pierwszy model (przewidyujący średnie wartości dzienne zanieczyszczeń) korzysta jedynie ze średnich danych dziennych dotyczących pyłu zawieszonego PM10. Analogiczne predykcje można wykonać dla pozostałych substancji, dla których istnieją pomiary. Być może informacje te korelowałyby ze sobą i dzięki temu model byłby jeszcze dokładniejszy.

Prognoza smogu mogłaby też pojawiać się przy najbliższych wydarzeniach jako kropka w 3 kolorach: zielonym, żółtym i czerwonym – intuicyjnie odpowiadającym jakości powietrza w chwili danego wydarzenia.

Jak można wdrożyć nasze rozwiązanie?

Jeśli chodzi o wydarzenia w Warszawie proponujemy dodanie do istniejącej już strony z kalendarzem (<https://um.warszawa.pl/kalendarz>) dodanie zaproponowanego przez nas rozwiązania z mapą oraz możliwości wyszukiwania po słowach kluczowych. Można by było także całą tę funkcjonalność umieścić w aplikacji Warszawa 19115.

W przypadku predykcji smogu, proponujemy dodanie prognozy smogowej na stronie <https://iot.warszawa.pl/#btn-air> oraz w aplikacji Warszawa 19115, gdzie znajdują się już aktualne informacje o jakości powietrza. Po kliknięciu na dany czujnik mogłoby się wyświetlać dodatkowo wartości przewidywane na kilka dni wprzód. Ciekawe byłoby także umożliwienie zobaczenia prognozy smogu dla całej Warszawy, np. wybierając w opcjach „jutro”.

Zbierając na bieżąco dane smogowe, proponujemy także wyjście poza ramy samej aplikacji, gdyż w momencie namnażania się różnorodnych rozwiązań, ludzie mogą poczuć się, jak w tzw. "aplikatopii". Z powodu zbyt dużej ilości aplikacji użytkownicy czują się tym przytłoczeni i nie korzystają z wielu z tych, które posiadają. Z tego względu proponujemy

wyjście z informacją smogową na ulicę, czyli na przykład prezentacja aktualnych danych o powietrzu oraz prognozy na kolejny dzień na ekranach/bilbordach. Dane te odpowiadałyby najbliższej stacji pomiaru jakości powietrza. Innym pomysłem jest ustawienie słupków świecących na te 3 kolory z opisem co przedstawiają. Można by byłoby umieścić takie światełka na słupkach już istniejących. Z pewnością takie niecodzienne rozwiązanie zaintryguowałoby one przechodniów i uświadamiało o istotności problemu i zainspirowało do zmian na lepsze. Dodatkowo aplikacja mogłaby wysyłać powiadomienia użytkownikom w momencie, gdy wartości zanieczyszczeń powietrza przekraczałyby pewne normy.