

Марија Јорданоска 203029
Тиана Костова 203061
Александра Пендовска 203089



Global Air Pollution

Анализа и предвидување на глобалното загадување на
воздухот во различни земји со примена на методи од машинско
учење





Вовед

Во текот на изминатите неколку децении, загадувањето на воздухот стана еден од најзначајните еколошки предизвици со кои се соочува светот. Зголемената урбанизација, индустријализацијата и транспортните активности доведоа до зголемени концентрации на штетни материји во воздухот, што има сериозни негативни влијанија врз јавното здравје и животната средина. Секојдневно, милиони луѓе ширум светот се изложени на високо ниво на загаденост на воздухот, што доведува до респираторни заболувања, кардиоваскуларни проблеми и прерана смрт.





Објаснување на проблемот

Еден од најголемите предизвици во решавањето на проблемот со загадувањето на воздухот е неговата сложеност и динамичност. Загадувањето на воздухот е резултат на комбинација од различни фактори кои можат да варираат во времето и просторот. Дополнително, интеракцијата помеѓу локалните и глобалните извори на загадување ја комплицира способноста за точно предвидување и управување со нивото на загадување.



Цел на проектот

Да се искористат напредните техники на машинско учење за да се анализираат нивоата на загаденост на воздухот во повеќе земји од светот и да се направи модел за предвидување на загаденоста на воздухот на главниот град на Индија, држава која придонесува најмногу за глобалниот Air Quality Index.

Анализа на глобалното загадување на воздухот во различни земји

Да се разбере влијанието на различните земји врз квалитетот на воздухот и да се спореди квалитетот на воздухот во различни земји во светот.

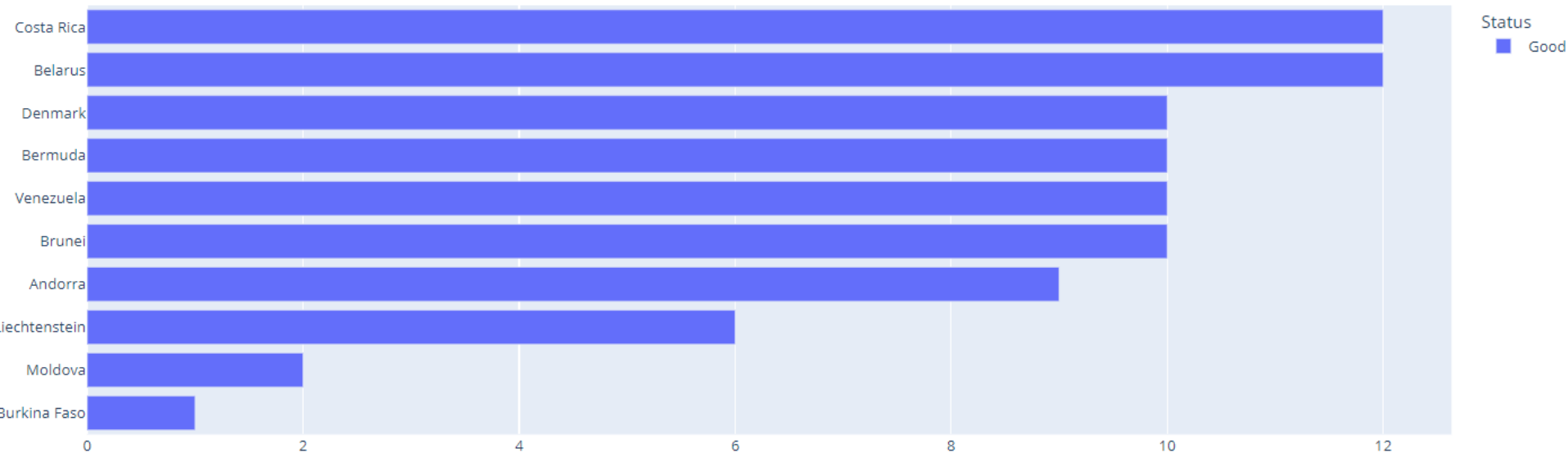


Air Quality Index



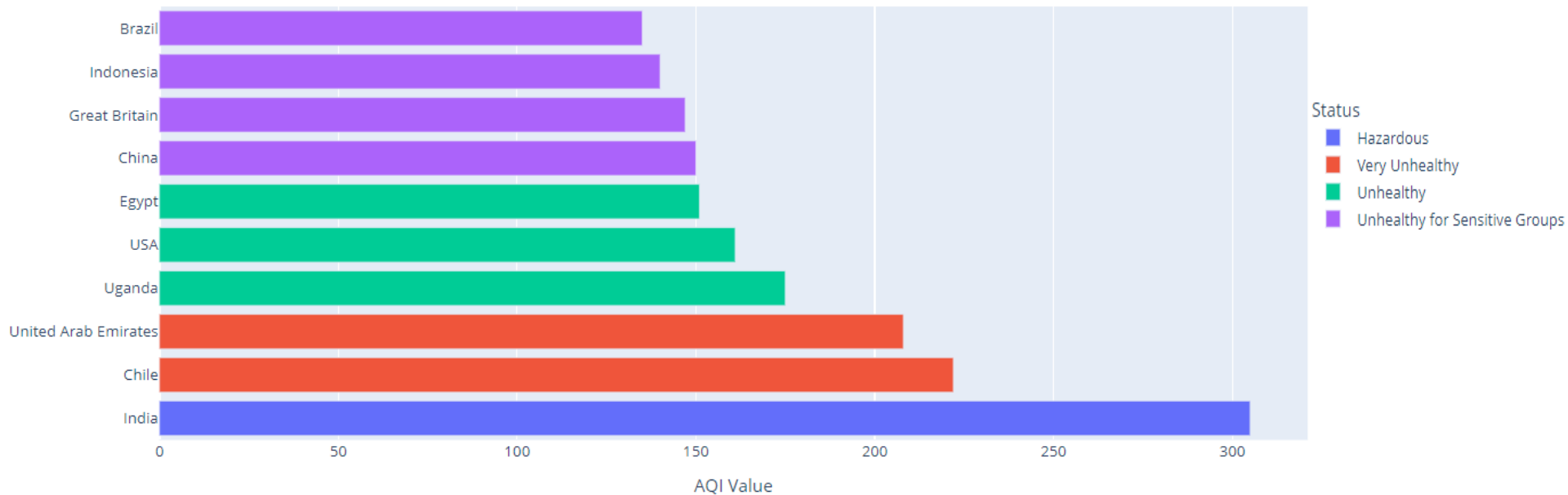


Топ 10 држави со најмал AQI на 04.07.2024





Топ 10 држави со највисок AQI на 04.07.2024



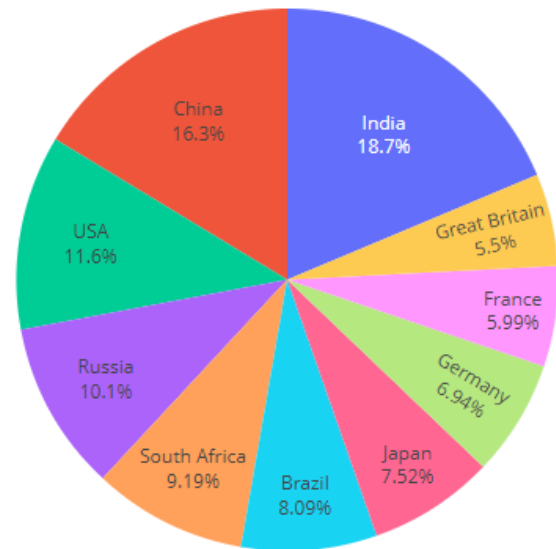
Глобален AQI во 2023 и 2024







Држави кои најмногу влијаат врз глобалниот AQI

- Според визуелизацијата, Индија е државата која најмногу влијае на глобалниот Air Quality Index.
- Го избираме нејзиниот главен град Њу Делхи за понатамошното истражување.





Градење на модел за предвидување на квалитетот на воздухот

Користејќи податочно множество од јавно достапна база, се фокусиравме на истражување и анализа на квалитетот на воздухот во Њу Делхи. Правевме предвидување на вредностите на AQI.



Фази на работа

1. Прибирање и подготовка на податоци
2. Развој на модел
3. Евалуација на модел
4. Резултати

1. Прибирање и подготовка на податоци

- Податоци превземени од јавно достапна база, со информации за мерењата на честичките кои влијаат врз AQI индексот

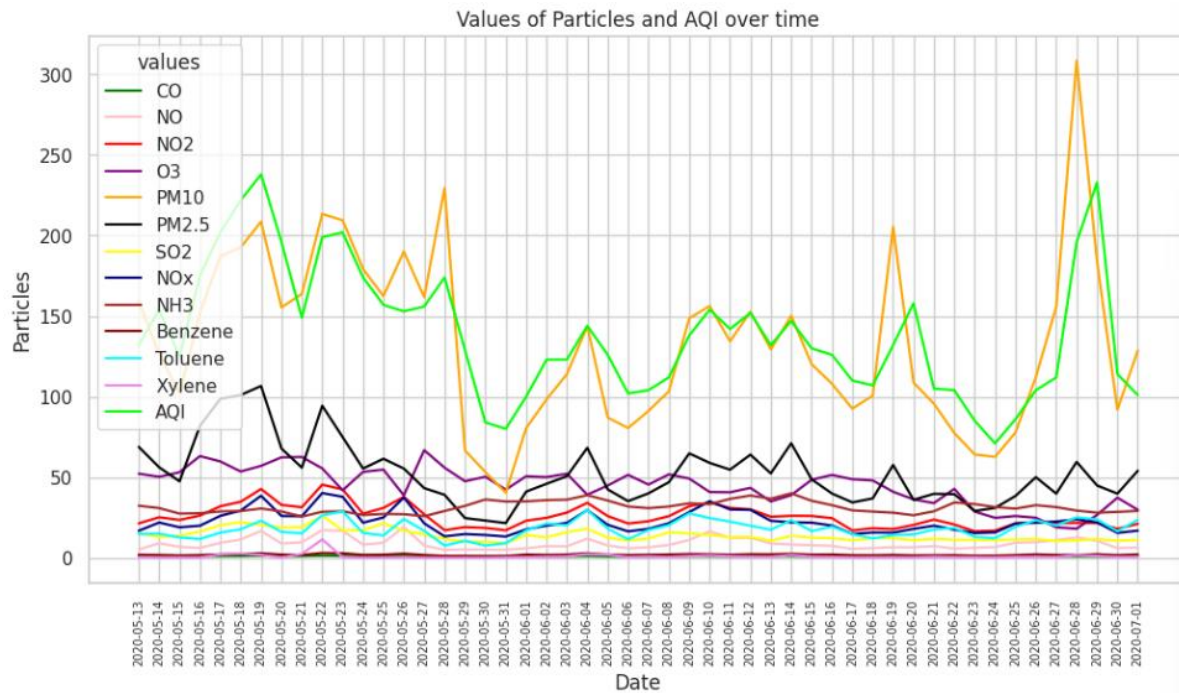
	City	Date	PM2.5	PM10	NO	NO2	NOx	NH3	CO	SO2	O3	Benzene	Toluene	Xylene	AQI	AQI_Bucket
10229	Delhi	2015-01-01	313.22	607.98	69.16	36.39	110.59	33.85	15.20	9.25	41.68	14.36	24.86	9.84	472.0	Severe
10230	Delhi	2015-01-02	186.18	269.55	62.09	32.87	88.14	31.83	9.54	6.65	29.97	10.55	20.09	4.29	454.0	Severe
10231	Delhi	2015-01-03	87.18	131.90	25.73	30.31	47.95	69.55	10.61	2.65	19.71	3.91	10.23	1.99	143.0	Moderate
10232	Delhi	2015-01-04	151.84	241.84	25.01	36.91	48.62	130.36	11.54	4.63	25.36	4.26	9.71	3.34	319.0	Very Poor
10233	Delhi	2015-01-05	146.60	219.13	14.01	34.92	38.25	122.88	9.20	3.33	23.20	2.80	6.21	2.96	325.0	Very Poor
...
12233	Delhi	2020-06-27	39.80	155.94	10.88	21.46	22.47	31.43	0.87	10.38	18.88	1.69	19.99	0.43	112.0	Moderate
12234	Delhi	2020-06-28	59.52	308.65	12.67	21.60	23.86	29.27	0.94	10.70	18.05	1.71	25.13	1.74	196.0	Moderate
12235	Delhi	2020-06-29	44.86	184.12	10.50	21.57	21.94	27.97	0.88	11.58	26.61	2.13	23.80	1.13	233.0	Poor
12236	Delhi	2020-06-30	39.80	91.98	5.99	17.96	15.44	28.48	0.84	10.51	37.29	1.57	16.37	0.49	114.0	Moderate
12237	Delhi	2020-07-01	54.01	128.66	6.33	21.05	16.81	29.06	0.97	11.15	29.73	2.03	23.57	0.65	101.0	Moderate

1. Прибирање и подготовка на податоци

- Справување со missing values со користење на KNN Imputer

```
from sklearn.impute import KNNImputer
imp=KNNImputer(n_neighbors=2)
data['PM2.5']=imp.fit_transform(data['PM2.5'].to_numpy().reshape(-1,1))
data['PM10']=imp.fit_transform(data['PM10'].to_numpy().reshape(-1,1))
data['NO']=imp.fit_transform(data['NO'].to_numpy().reshape(-1,1))
data['NO2']=imp.fit_transform(data['NO2'].to_numpy().reshape(-1,1))
data['NH3']=imp.fit_transform(data['NH3'].to_numpy().reshape(-1,1))
data['SO2']=imp.fit_transform(data['SO2'].to_numpy().reshape(-1,1))
data['O3']=imp.fit_transform(data['O3'].to_numpy().reshape(-1,1))
data['Xylene']=imp.fit_transform(data['Xylene'].to_numpy().reshape(-1,1))
data['AQI']=imp.fit_transform(data['AQI'].to_numpy().reshape(-1,1))
```

1. Прибирање и подготовка на податоци



2. Развој на модел

- Делење на множеството на тренирачко и тестирачко со `train_test_split` функцијата
 - Во тренирачкото множество (X) се сите влезни податоци на моделот, колоните со вредностите за сите честички освен AQI
 - Во тестирачкото множество (Y) е таргет колоната AQI

2. Развој на модел

LSTM Neural Network со три слоја:

```
model.add(LSTM(64, input_shape=(1,12), activation='relu', return_sequences=True))  
  
model.add(LSTM(64, activation='relu'))  
  
model.add(Dense(1 ,activation='linear'))
```

2. Развој на модел

Компајлирање на моделот:

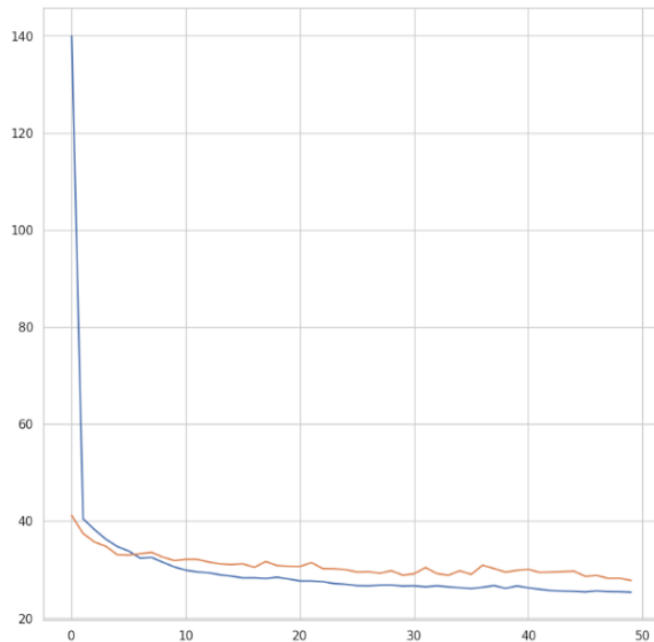
```
model.compile(optimizer='adam', loss='mean_squared_error', metrics=['mean_absolute_error'])
```

Тренирање на моделот:

```
history = model.fit(x_train, y_train, shuffle=False, validation_split=0.2, epochs=50, batch_size=16)
```

3. Евалуација на моделот

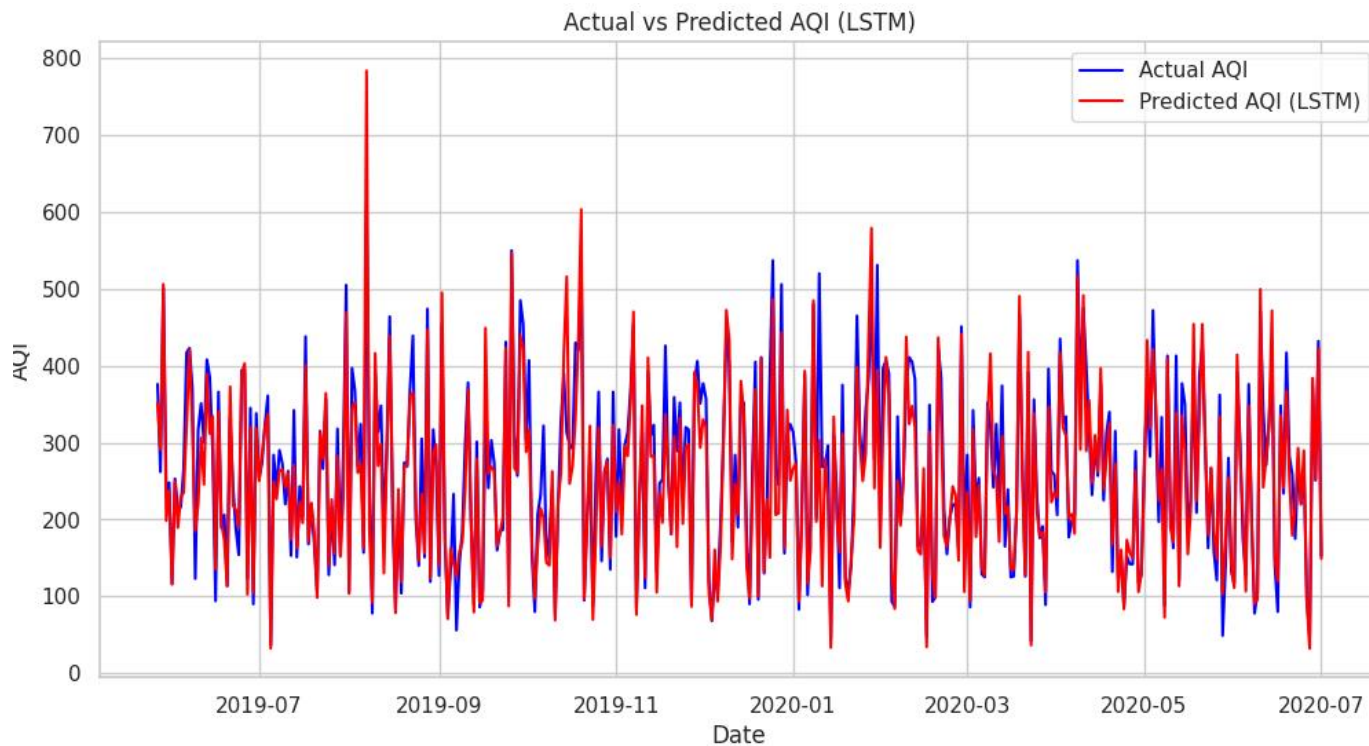
Train и validation за секоја епоха:



3. Евалуација на моделот

- R^2 score = **88%**
- Mean Absolute Percentage Error = **11%**


4. Результати





Заклучок

Ова истражување за анализа и предвидување на загаденоста на воздухот покажа дека Индија е земјата која најмногу влијае врз глобалното загадување на воздухот. Според тоа, направен е модел за предвидување на AQI вредностите во Њу Делхи. Податочното множество искористено во ова истражување има информации за вредностите на различните честички измерени во воздухот во период од 2015 до 2020 година. На податоците е направена подготовка и визуелизација. Изграден е модел на LSTM невронска мрежа која ги предвидува AQI вредностите. Моделот има 88% точност. Овој модел може да служи како основа за понатамошни истражувања и развој на поефективни мерки за контрола на загадувањето



**Ви благодариме на
вниманието!**

