1. Vizualizacija podatkov in ocenjevanje atributov
   1. Overview of the initial data set and added attribues

The initial dataset consists of the 2478 observations and 27 variables. None of the variable contain missing data.

During the analysis and estimation of the initial data set 11 more attributes were added to the data set. Description of the attributes and formulas to obtain the new values are presented in the table below.

|  |  |  |
| --- | --- | --- |
| **Atribut** | **Class** | **Description** |
| Datum | factor | Date value provided in the initial dataset |
| Postaja | factor | Postaja values provided in the initial dataset  Possible option: Koper, Ljubljana |
| Glob\_sevanje\_max | numeric | Numeric value provided in the initial dataset |
| Glob\_sevanje\_mean | numeric | Numeric value provided in the initial dataset |
| Glob\_sevanje\_min | numeric | Numeric value provided in the initial dataset |
| Hitrost\_vetra\_max | numeric | Numeric value provided in the initial dataset |
| Hitrost\_vetra\_mean | numeric | Numeric value provided in the initial dataset |
| Hitrost\_vetra\_min | numeric | Numeric value provided in the initial dataset |
| Sunki\_vetra\_max | numeric | Numeric value provided in the initial dataset |
| Sunki\_vetra\_mean | numeric | Numeric value provided in the initial dataset |
| Sunki\_vetra\_min | numeric | Numeric value provided in the initial dataset |
| Padavine\_mean | numeric | Numeric value provided in the initial dataset |
| Padavine\_sum | numeric | Numeric value provided in the initial dataset |
| Pritisk\_max | numeric | Numeric value provided in the initial dataset |
| Pritisk\_mean | numeric | Numeric value provided in the initial dataset |
| Pritisk\_min | numeric | Numeric value provided in the initial dataset |
| Vlaga\_max | numeric | Numeric value provided in the initial dataset |
| Vlaga\_mean | numeric | Numeric value provided in the initial dataset |
| Vlaga\_min | numeric | Numeric value provided in the initial dataset |
| Temperatura\_Krvavec\_max | numeric | Numeric value provided in the initial dataset |
| Temperatura\_Krvavec\_mean | numeric | Numeric value provided in the initial dataset |
| Temperatura\_Krvavec\_min | numeric | Numeric value provided in the initial dataset |
| Temperatura\_lokacija\_max | numeric | Numeric value provided in the initial dataset |
| Temperatura\_lokacija\_mean | numeric | Numeric value provided in the initial dataset |
| Temperatura\_lokacija\_min | numeric | Numeric value provided in the initial dataset |
| PM10 | numeric | Numeric value provided in the initial dataset |
| O3 | numeric | Numeric value provided in the initial dataset |
| PM10\_Class | factor | Added variable obtained based on the values of the PM10 Možni razredi so: NIZKA (do 35.0), VISOKA (nad 35.0) |
| O3\_Class | factor | Added variable obtained based on the values of the O3 Možni razredi so: NIZKA (pod 60.0), SREDNJA (med 60.0 in 120.0), VISOKA (med 120.0 in 180.0) in EKSTREMNA (nad 180.0) |
| leto | integer | Added variable representing the year data was collected obtained from extracting the year value from the variable Datum |
| mesec | integer | Added variable representing the month data was collected obtained from extracting the year value from the variable Datum |
| Mesec\_Abb | factor | Added variable representing the abbrevation of the month data was collected obtained from extracting the year value from the variable Datum |
| letni\_cas | character | Added variable representing the season data was collected obtained from extracting the year value from the variable Datum |
| Glob\_sevanje\_spr | numeric | Added variable representing the change of the Globalno sevanje, calculated as: Glob\_sevanje\_max - Glob\_sevanje\_min |
| Pritisk\_spr | numeric | Added variable representing the change of the Pritisk, calculated as: Pritisk\_max - Pritisk\_min |
| Vlaga\_spr | numeric | Added variable representing the change of the Vlaga, calculated as: Vlaga\_max - Vlaga\_min |
| Temperatura\_Krvavec\_spr | numeric | Added variable representing the change of the Temperatura\_Krvavec, calculated as: Temperatura\_Krvavec\_max - Temperatura\_Krvavec\_min |
| Temperatura\_lokacija\_spr | numeric | Added variable representing the change of the Pritisk, calculated as: Temperatura\_lokacija\_max - Temperatura\_lokacija\_min |

* 1. Summary statistics for the initial dataset

Pripravite nekaj zanimivih grafov, ki ilustrirajo podane podatke (distribucije vrednosti, soodvisnosti med atributi, ponavljajoče se vzorce in podobno).

2. Ocenjevanje atributov Ocenite kvaliteto podanih atributov in konstruirajte nove atribute, ki lahko izboljšajo kvaliteto zgrajenih modelov. Namig: datum je v obstoječi obliki relativno neuporaben, iz njega pa lahko izpeljemo nove atribute (npr. letni čas, dan v tednu…), ki potencialno pomagajo pri napovedovanju onesnaženja zraka.

kreiranje vsaj enega dodatnega atributa; pri gradnji modelov poskušajte izbrati ustrezno podmnožico atributov in primerjajte tako dobljene modele s tistimi, ki so naučeni na celotni učni množici.

3. Klasifikacija Zgradite vsaj tri klasifikacijske modele za napovedovanje: a. maksimalne dnevne koncentracije ozona – možni razredi so: NIZKA (pod 60.0), SREDNJA (med 60.0 in 120.0), VISOKA (med 120.0 in 180.0) in EKSTREMNA (nad 180.0), b. dnevne koncentracije delcev PM10 – možni razredi so: NIZKA (do 35.0), VISOKA (nad 35.0).

4. Regresija Zgradite vsaj tri regresijske modele za napovedovanje: a. maksimalne dnevne koncentracije ozona, b. dnevne koncentracije delcev PM10.

5. Evalvacija modelov Zgrajene modele ustrezno ovrednotite in predstavite dobljene rezultate. Primerjajte različne modele na podlagi dosežene točnosti napovedovanja in razumljivosti delovanja.

6. Poročilo (dokument v formatu .doc ali .pdf) V poročilu opišite svoj pristop, uporabljene modele in atribute, predstavite dosežene rezultate ter strnite zaključke na podlagi eksperimentalne evalvacije.