

|  |
| --- |
| EPİDEMAI (ENG)  2023  Proje Kapsamı |
|  |
| 9 Ağustos  By Documentation: Pınar Topuz |

# Project Name: Modeling and Prediction of Disease Outbreaks

|  |
| --- |
| Aim:  This project is a data awareness project that aims to analyze disease outbreaks and predict future trends. The main purpose of our project is to collect, analyze and visualize data on disease outbreaks and to predict the future course of epidemics using epidemiological models. These projections will provide guidance on planning health services, taking emergency measures and optimizing resource allocation. |
|  |
| Technologies Used Programming Language::   * Python Data Analysis Libraries: Pandas, NumPy * Data Visualization Libraries: Matplotlib, Seaborn * Machine Learning Libraries: Scikit-learn, TensorFlow   Project Structure:  Step 1: Data Collection and Preparation  We identified data sources: Health organizations, World Health Organization, etc. We collected and cleaned datasets: Filled in missing data, outliers will be addressed.  Step 2: Data Visualization  We analyzed the basic statistics: The mean, median, standard deviation will be calculated. We did a time series analysis: The number of cases and deaths will be visualized with a time graph. We created geographic visualizations: It will show the geographical distribution of the epidemic with maps.  Step 3: Model Development  Epidemiological Model Selection: The SEIR model has been selected and the equations will be determined. Data-Based Parameter Estimates: Model parameters will be adapted to the data. Machine Learning and Artificial Intelligence: Prediction models will be developed using machine learning algorithms.  Step 4: Evaluation of Model Evaluation and Improvement Estimation Performance:  Model performance will be measured with error metrics. Model Improvement: Parameter adjustments and alternative algorithms will be tried.  Step 5: Reporting and Presentation  Detailed Documentation: Jupyter Notebook will be documented with content and step-by-step explanations. Support with Visuals and Diagrams: Images and diagrams will be added to support the project.  Step 6: Presentation Preparation  Creating Effective Presentations: An effective presentation will be prepared for technology organizations. Answering Questions: You will be prepared for the questions that may come during the presentation.  Step 3: Model Development  Epidemiological Model Selection: The SEIR model has been selected and the equations will be determined. Data-Based Parameter Estimates: Model parameters will be adapted to the data. Machine Learning and Artificial Intelligence: Prediction models will be developed using machine learning algorithms.  Step 4: Evaluation of Model Evaluation and Improvement Estimation Performance:  Model performance will be measured with error metrics. Model Improvement: Parameter adjustments and alternative algorithms will be tried.  Step 5: Reporting and Presentation  Detailed Documentation: Jupyter Notebook will be documented with content and step-by-step explanations. Support with Visuals and Diagrams: Images and diagrams will be added to support the project.  Step 6: Presentation Preparation  Creating Effective Presentations: An effective presentation will be prepared for technology organizations. Answering Questions: You will be prepared for the questions that may come during the presentation. |

Project Diagram:

[Project Start]

|

v

[Data Collection and Preparation]

|

v

[Data analysis]

|

v

[Data Visualization]

|

v

[Model Development]

|

v

[Model Evaluation and Refinement]

|

v

[Reporting and Presentation]

|

v

[Project Completion]

Models and Methods Used:

Epidemiological Model: SEIR Model

Model Description: The SEIR model (Suscetible, Exposed, Infected, Recovered) is an epidemiological model that mathematically models the spread of a disease. In this model, population members are divided into four categories: susceptible, exposed, infected, and recovered. The model represents the transitions between these categories with systems of equations.

Model Equations:

Number of People with Prayers:

�

�

�

�

=

−

�

⋅

�

�

�

dt

dS

=−β⋅

N

SI

Number of Persons in the Incubation Period:

�

�

�

�

=

�

⋅

�

�

�

−

�

⋅

�

dt

also

=β⋅

N

SI

−σ⋅E

Number of Infected Persons:

�

�

�

�

=

�

⋅

�

−

�

⋅

�

dt

dI

=σ⋅E−γ⋅I

Number of Recovered Persons:

�

�

�

�

=

�

⋅

�

dt

dR

=γ⋅I

Here,

�

S-sensitive people,

�

E incubation period,

�

I infected persons,

�

R represents the number of people who have recovered.

�

β is the contact velocity,

�

σ is the incubation period pass rate,

�

γ is the recovery pass rate, and

�

N stands for the total population.

Machine Learning Algorithms Used

Model Selection: Apart from the epidemiological model, we analyzed the data using machine learning algorithms such as linear regression, decision trees, and random forest. Because these algorithms have different learning capabilities, we developed different prediction models.

Project Progress

Work-Timeline

Literature Review: April - May

We reviewed relevant articles, researched existing models.

Data Collection and Preparation: May - June

We collected data from health institutions and filled in missing data.

Data Analysis and Visualization: June - July

We performed basic statistics and time series analysis.

We visualized the geographical distribution with maps.

Model Development: July - August

We started to implement the SEIR model and machine learning algorithms.

We adapted the model parameters to the data.

Model Evaluation and Improvement: August - September

We evaluated model performance, calculated error metrics.

We made parameter adjustments and model improvements.

Reporting and Presentation: September - October

We created the Jupyter Notebook content, added images and diagrams.

We started preparing an effective presentation.

Findings and Conclusions

As a result of the project, data analysis and epidemiological modelling, we observed:

We were able to identify trends regarding the rate of spread of epidemics and the peak period.

We analyzed how parameters such as the rate of recovery and the number of infected people affect the course of the epidemic.

We evaluated how machine learning algorithms performed relative to epidemiological models.

Recommendations and Progress

More data collection can increase the reliability of the estimates, especially for different regions and subgroups.

More complex variations of the model can be studied to better understand the different interactions between individuals.

By integrating real-time data flow, we can predict outbreaks faster.

Source

[Reference Article 1]

[Reference Article 2]

[Data Source Used]

[Machine Learning Book]

Appendix and Appendages

Epidemiological model code fragments and equations

Python codes using machine learning algorithms

Full-size versions of related images and graphics

Project-related correspondence or communication samples