Nature-Inspired Computing Project Report

Ayhem Bouabid
DS-01
Innopolis University
Innopolis, Russian Federation
a.bouabid@innopolis.university

Majid Naser
DS-01
Innopolis University
Innopolis, Russian Federation
m.naser@innopolis.university

Nikolay Pavlenko
DS-01
Innopolis University
Innopolis, Russian Federation
n.pavlenko@innopolis.university

***Abstract*—tbd**
***Index Terms*—tbd**

I. Introduction

Quality data preprocessing is an important prerequisite to creating accurate machine learning models. It solves a wide range of problems, from missing data and data inconsistency to required anonymization. In our project we have decided to focus primarily on one aspect of it, that is, **feature selection**.

Performing feature selection allows us to solve a great number of problems, being especially effective in large datasets containing many features. It improves accuracy of predictions by finding and eliminating spurious relationships, as well as reducing the chances of overfitting. Other effects of feature selection are the improvement of training time for the model through cutting down on unnecessary data, and increase in interpretability, as fewer features have to be analyzed to figure out the dependencies.

However, the main problem that can be solved by feature selection for multiple regression models (and will be addressed specifically in our project) is **multicollinearity**. Multicollinearity is an effect when multiple explanatory variables that are believed to me independent from each other, are in fact, closely interrelated. This can have damaging effects on the accuracy of prediction models, as even a small change in data will lead to unpredictable results. Ideally, feature selections would find such relationships and purge the dataset from them, our model aims to do so as well.

As to the practical applications of our project, we have decided to stick to the original project proposal and test it on a new dataset that would contain various countrywide statistics as features that will try to predict the population growth for a country in a given year. Since we have found no such dataset that would have suited our needs, it was created from scratch, using data provided by the World Bank Open Data site [1]. However, during our work on the project, we have decided to also test other similar datasets, that could be used for a multiple regression model, to figure out if our implemented feature selection process can be used more generally.

NEW INFO ABOUT OTHER DATASETS WILL BE ADDED HERE

II. Related Work

tbd

III. Methodology

In our research, we were able to identify three categories of feature selection methods. Firstly, it is the filter method, which ranks each feature on some statistical metric, and evaluates the ranks afterwards, picking the ones that score the highest. Secondly, it is the wrapper method, which takes a subset of features and trains a model using them. Depending on results of the testing, it adds or removes features from the subset, incrementally improving the performance until user-defined stopping criteria is achieved. Thirdly, it is the embedded method, which is in-built into models themselves - it add a penalizing term to the regression equation (en example of such a model would be Lasso Regression).

Among all categories mentioned, wrapper method has the highest computational costs, but can provide the best dataset that would provide the most accurate results for our model. It can also include nature-inspired algorithms as its estimators, making **wrapped** category of feature selection models our preferred choice.

IV. Github link

$https://github.com/Daru1914/NIC\_Project$

V. Experiments and Evaluation

tbd

VI. Analysis and Observations

tbd

VII. Conclusion

tbd

VIII. References

References

[1] The World Bank. World bank open data. https://data.worldbank.org/. Accessed: 20122-12-02.

TABLE I
TABLE TYPE STYLES

Table Head	**Table Column Head**		
	Table column subhead	***Subhead***	***Subhead***
copy	More table copy[a]		

[a]Sample of a Table footnote.

IX. FUNNI TABLE

X. FUNNI PICTURE

Fig. 1. Example of a figure caption.