# Table of Content

[Table of Content 1](#__RefHeading___Toc2691_1970215514)

[List of Figures 2](#__RefHeading___Toc2485_1970215514)

[List of Tables 3](#__RefHeading___Toc3012_1970215514)

[Acknowledgments 4](#__RefHeading___Toc2910_1970215514)

[Dedication 5](#__RefHeading___Toc2707_1970215514)

[Abstract 6](#__RefHeading___Toc2701_1970215514)

[Introduction 7](#__RefHeading___Toc3010_1970215514)

[Chapter 1: State of the art 8](#__RefHeading___Toc2699_1970215514)

[1. Machine learning 8](#__RefHeading___Toc249_2222239152)

[2. Class decomposition 8](#__RefHeading___Toc216_1432413569)

[Chapter 2: Conception 9](#__RefHeading___Toc2697_1970215514)

[Chapter 3: Experimentation and Results 10](#__RefHeading___Toc2695_1970215514)

[Conclusion and Perspectives 11](#__RefHeading___Toc3017_1970215514)

[References 12](#__RefHeading___Toc2693_1970215514)

[Bibliography 13](#__RefHeading___Toc3015_1970215514)

# List of **F**igures

[Figure 1: Class decomposition 9](#Figure!0|sequence)

# List of Tables

...

# Acknowledgments

I thank above all the god who gave me the courage and the will to complete this work.

I deeply thank my supervisor Mrs. Seridi for guiding me with patience, and for all her efforts, advice and corrections.

My heartfelt thanks to my friends who have helped me.

Finally, a big thanks to my family for supporting me during my university course.

# Dedication

To the whole family

To all my friends

# Abstract

...

# Introduction

Flood forecasting is the process of assessing and predicting the volume, timing, and length of floods based on known features of a river basin, with the goal of preventing harm to people, property, and the environment. it is required for establishing effective flood risk management strategies, reducing flood danger, evacuating people from flood-prone locations, and managing water resources systems. In recent years, data-driven techniques/models for flood forecasting have gotten a lot of attention, incorporating modern machine learning techniques and algorithms like artificial neural network, decision trees, random forest, etc.

In this work we aim to create a decision support system with modern machine learning techniques by the classification of flood zones while implementing class decomposition to extract important insight from the datasets.

# Chapter 1: State of the art

## Flood forecasting

Flood forecasting is the act of assessing and predicting the volume, timing, and length of floods based on known aspects of a river basin in order to avoid harm to people, property, and the environment. It's necessary for developing effective flood risk management plans, minimizing flood hazard and evacuating people from flood-prone areas.

## Machine learning

Machine learning, also known as artificial learning, is a form of artificial intelligence (AI) that allows a system to learn from data and not through explicit programming. As algorithms ingest training data, it becomes possible to create more accurate models based on this data. A machine learning model is the result generated when you train your machine learning algorithm with data. After the training, when you provide input to a model, you receive an output result. a predictive algorithm creates a predictive model, when data is provided to the predictive model, it results in a forecast that is determined by the data that formed the model.

## Artificial Neural Networks

An Artificial Neural Network (ANN) is a data or signal processing system made up of a large number of simple processing pieces linked together by direct linkages to execute parallel distributed processing in order to complete a specific computational goal. Neural networks act in a similar fashion to the human brain when it comes to processing information. Neural networks learn via example, similar to how biological nerve systems, such as the brain work.  
In comparison to traditional computing, ANN takes a distinct approach to problem solving. In order to solve a problem, traditional computer systems employ an algorithmic method, which entails following a series of instructions. This restricts our problem-solving abilities to problems that we are familiar with and can solve. Neural networks and traditional algorithmic computers, on the other hand, do not compete but rather complement each other. There are activities that are better suited to an algorithmic method, such as arithmetic operations, and tasks that are better suited to a neural network approach, such as image processing.

## **Decision Tree**

A decision tree is a very simple model. Given several characteristics, the decision begins with one of these characteristics; if this is not enough, another one is used, and so on. It is widely known and used in many companies to facilitate the decision-making process and risk analysis. It was widely used in the 1960s and 1980s for the construction of expert systems. The rules are introduced manually, for this reason this model lost its popularity after the 80s. The emergence of mathematical methods to build decision trees has brought this model back to the battle of algorithms of automatic apparent drawing.

## Class decomposition

The technique of segmenting each class into a number of homogeneous sub-classes is known as class decomposition. Clustering is a natural way to accomplish this. Using class decomposition in supervised learning, particularly ensembles, can provide a variety of advantages. It can be a computationally efficient way to generate a linearly separable dataset without the requirement for feature engineering, which is required by techniques such as Support Vector Machines (SVM) and Deep Learning. Decomposition is a natural technique for ensembles to promote diversity, which is a fundamental aspect in the effectiveness of ensemble classifiers.

## **Problem**

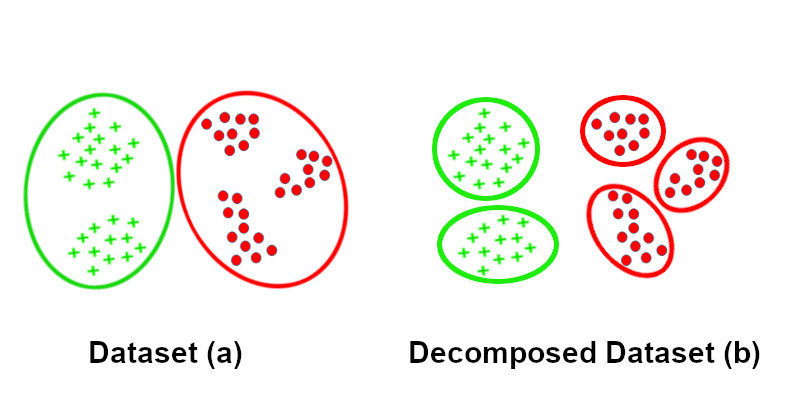


Figure 1: Class decomposition

Many datasets, including the ones we used in this work, are binary datasets that have only two classes which are represented by the values: 0, False or 1, True. They are created after some classification is applied to the data, and thus in the process of predicting, which is called binary classification, the output is restricted to only two classes, but in some cases having more diverse classes in the dataset helps our model better understand and classify our data by learning the more in-depth patterns.

Having the output restricted to only two classes, we may miss important insight that might be found in our data, like sub-classes in medical datasets might represent a sub-type of a disease, knowing which sub-type of a disease a patient has helps doctors give a more precise and effective treatment to their patients.

## **Decomposition**

Class decomposition facilitates learning class boundaries of a dataset for a machine learning model and consequently may improve the precision of the given model, as well as uncovering valuable insight from the dataset, there many ways of achieving class decomposition

## **Related works**

In 2019 Samih M. Mostafa and Hirofumi Amano 1 published an article where they highlighted the positive effect that clustering has on machine learning models accuracy, their work presents a technique where the data is clustered using the K-means algorithm, the number of clusters was determined by the elbow method which is a heuristic method of validation and interpretation of symmetry within cluster analysis, in this method the clustering processing step was done before applying the prediction algorithm.

In the experimentation phase the authors used MLR, Ridge, Lasso and ElasticNet machine learning algorithms and the results showed that the proposed method achieves significant improvement from the point of view of RMSE, and coefficient of determination 𝑅². The work done by these individuals that there are potential benefits to data clustering before applying a prediction algorithm.

Apart from the possible performance benefits that class decomposition may offer to a machine learning model, there are other benefits that can be that can be gained that are related to the insight we can extract from our datasets, the work done by Suchi Saria, Anna Goldenberg in 2015 2 highlights the problem with binary medical datasets and that some additional important information can be extracted from them that can helps with doctors decision making and treatment choices, this approach is called precision medicine, which is the main focus of this article, the discovery and refinement of disease sub-types can benefit both the practice and science of medicine. Clinically, by refining prognoses based on similar individuals, disease sub-types help reduce uncertainty in an individual’s expected outcome. Accurate prognoses can thereby improve treatment decisions.

They showed examples of binary datasets of different diseases where decompositing the positive class into multiple classes using statistical and machine learning approaches such as nonnegative matrix factorization, hierarchical clustering, and probabilistic latent factor analysis, reveals the presence of sub-types related to that disease.

## **Genetic algorithm**

## **Conclution**

# Chapter 2: Conception

## **Introduction**

In this work we aim

# Chapter 3: Experimentation and Results

…

# Conclusion and Perspectives

…

# References

[1] Samih M. Mostafa, Hirofumi Amano, Effect of clustering data in improving machine learning model accuracy, 2019, https://kyushu-u.pure.elsevier.com/en/publications/effect-of-clustering-data-in-improving-machine-learning-model-acc