With the support of the Erasmus+ Programme of the European Union

EMJMD GENIAL

GREEN NETWORKING & CLOUD COMPUTING

LEEDS BECKETT UNIVERSITY

The Italian scenario of solid waste: composition, recycling rates and segregation.

Author:	Marianna Oleotti
Module:	Data Analysis and Visualization
Supervisor:	Dr. Ah-Lian Kor
Faculty:	School of Built Environment, Engineering, and Computing
Course:	MSC Information & Technology - Sustainable Computing
Level:	Level 7
Date:	05/2022

Abstract

Waste management represents a pervasive challenge in nowadays society, as a result of the constant population growth, and the consumerist lifestyle. Earth is a closed system: the resources and the amount of matter on the planet are finite. Therefore, the need to switch towards a Circular Economy is of utmost importance, in a world that has been discarding material in a Linear Economy trend for the past centuries. The 3R Initiative proposed by the G8 Summit of 2004 promotes the "3Rs" (reduce, reuse and recycle) through the effective use of resources and materials.

This report looked in depth into the waste production and recycling of Italy, one of the countries of the European Union with the highest rate of waste sorting. The dataset was first explored through Descriptive and Inferential Statistic to identify the least performant areas of Italy, and relationship between the production and recycling rates; then, through Machine Learning forecasting algorithms, the values are predicted for the next years to understand the future amount of waste produced by the country. As one the first step for a better recycling starts with waste sorting in citizens everyday life, Deep Learning lightweight classification neural networks (MobileNet and DenseNet121) are applied to a dataset of trash images in the context of Smart Bins for Automated Waste Sorting. Evidence-based recommendations are given at the end of the analysis.

Table of contents

Abstract **i**

Table of contents **ii**

List of figures **iv**

List of tables **vi**

1 Introduction **1**
1.1 Background 1
1.2 Aim and Objectives 2
1.2.1 RO1 – Level 1 2
1.2.2 RO2 – Level 2 2
1.2.3 RO3 – Level 3 3
1.2.4 RO4 – Level 4 3
1.3 Rationale 3
1.4 Scoping of Research 4
1.5 Contribution to Research 4
1.6 Organization of the Report 5

2 Literature Review **6**
2.1 Policies and regulations of solid waste management 8
2.2 Waste segregation 11
2.2.1 Smart Solutions 12

3 Methodology **15**
3.1 Macro methodology 15
3.1.1 Data collection 15
3.1.2 Data dictionary 16
3.1.3 Data pre-processing 18

3.2 Micro methodology . . . 19

4 Results and Discussion **22**
4.1 EDA: first level of analysis . . . 22
4.1.1 Recycling rates . . . 22
4.1.2 Waste production . . . 25
4.1.3 Waste composition . . . 26
4.2 Inferential Statistic: second level of analysis . . . 27
4.2.1 Solid waste composition . . . 27
4.2.2 Correlation . . . 30
4.3 Machine Learning: third level of analysis . . . 31
4.3.1 Waste production . . . 31
4.3.2 Recycling rates . . . 35
4.4 Deep Learning: fourth level of analysis . . . 41

5 Recommendations **43**

6 Conclusion and Future Works **44**
6.1 Conclusion . . . 44
6.2 Future Work . . . 44

References **52**

A Appendix **53**
A.1 Methodology . . . 53
A.2 Results . . . 56
A.3 Descriptive statistics . . . 56
A.3.1 Recycling rates . . . 56
A.3.2 Waste production . . . 60
A.3.3 Waste composition . . . 62
A.4 Inferential Statistic . . . 63
A.5 Machine Learning . . . 66
A.6 Deep Learning . . . 68

List of Figures

1 EU Waste Hierarchy 8
2 A comparison of the color scheme for waste collection and sorting before UNI standard 12
3 Dataset missing values 18
4 Structure of MobileNet (v1 20
5 Structure of DenseNet121 21
6 Head of the dataset used for the analysis 22
7 Percentage of solid waste recycled in Italy in 2020 23
8 Percentages of solid waste recycled in regions of Italy in 2020 23
9 Distribution of municipalities' recycling rate in Liguria 24
10 Percentages of solid waste recycled in the provinces of Liguria 24
11 Map of the recycling rates of the municipalities of Liguria 25
12 National recycable and unsorted solid waste per capita in kg 26
13 Regional recycable and unsorted solid waste per capita in kg 26
14 Macro-regional unsorted solid waste per capita in kg 27
15 Recycable solid waste composition 27
16 Distribution of waste composition over macro-regions 28
17 Result of the ANOVA test performed on wood waste 29
18 Result of the ANOVA test performed on glass waste 29
19 Result of the ANOVA test performed on paper waste 29
20 Result of the ANOVA test performed on organic waste 29
21 Regional percentages of wood waste production (over the total amount of waste) 30
22 Heatmap of correlation among attributes 31
23 Regional GDP of Italy in 2020 32
24 National waste production per capita trend 32
25 Model coefficients and errors 33
26 Macro-regional waste production per capita future trend 33
27 Model coefficients and errors for North Italy 34

28 Model coefficients and errors for Center Italy 34
29 Model coefficients and errors for South Italy 34
30 Model coefficients and errors for Region Liguria 34
31 Waste production per capita trend for Region Liguria 35
32 Model coefficients and errors for Region Molise 35
33 Waste production per capita trend for Region Molise 36
34 National waste production per capita trend 37
35 National waste production per capita trend 37
36 National waste production per capita trend 38
37 Macro-regional recycling rate future trend 38
38 Model coefficients and errors for North Italy 39
39 Model coefficients and errors for Center Italy 39
40 Model coefficients and errors for South Italy 39
41 Recycling rate and errors for Region Veneto 39
42 Recycling rate trend for Region Veneto 39
43 Recycling rate and errors for Region Sicily 40
44 Recycling rate trend for Region Sicily 40
45 Images per class of the Garbage Collection dataset 41
46 MobileNet model 42
47 DenseNet121 model 42

List of Tables

1 Typical composition of municipal solid waste 7
2 Examples of apps available on the market 12
3 List of the use of CNN for waste segregation 14
4 Parameters tested on the models 41

Chapter 1

Introduction

1.1 Background

In the era of consumerism and fast fashion, 2.01 billion tonnes of municipal solid waste are generated every year. According to Statista, this value is expected to reach 3.4 billion by 2050 due to urbanization, economic and population growth.

This is becoming a significant concern, since only 15% percent of it is recycled or reused, as reported by the Organisation for Economic Co-operation and Development (OECD). The remaining is dumped in landfills, causing the 2% of the global greenhouse gas emissions of 2016 (Ritchie et al. 2020). With the Circular Economy Action Plan of 2012, the European Union and the State Members commit to adopt a model that “gives back to the planet more than it takes” (European Commission 2020), switching to consumption patterns that reintegrate as much secondary raw materials as possible. The 3R Initiative is the core of the circular economy: reduce, recycle and reuse; but the problem of recycling and reuse has been addressed since the 1970s by a number of OECD countries.

Therefore, it is of utmost importance to focus on the first step of better material recycling: household waste sorting. As a matter of fact, inadequate sorting or lack of sorting is one of the biggest issues in waste management and it contributes to pollution and natural resource depletion (Minelgaitė & Liobikienė 2019).

In this report, Italian waste production, composition and sorting is analyzed as a case study of a standardized investigation that should be performed by all countries to identify the interventions still needed to achieve EU goals.

As outlined by the study of Minelgaitė & Liobikienė (2019) one of the reasons behind the not waste sorting behaviour is the missing trust in the recycling system or the inconvenience of doing it.

The automatic waste sorting in households is the consequent solution to this problem as it could help citizen improve their recycling accuracy and, at the same time, the automatized process could prove the efficiency of the waste segregation process.

This report will focus also on the study of two Deep Learning algortihms to perform waste sorting, implementable in Smart Bins.

1.2 Aim and Objectives

The aim of this report is to analyze the trends of municipal solid waste production and sorting in the Italian scenario, and inquire on how these trends differ. The objectives can be divided into four chapters, corresponding to the four levels of analysis, which will be further detailed in the below subsections.

1.2.1 RO1 – Level 1

An Exploratory Data Analysis (EDA) will be performed to study the production of solid waste at a regional, macro-regional and national level. This analysis includes multiple parameters such as mean, median, minimum, maximum, standard deviation and quartiles, and visualizations such as box plots, bar charts, histograms, maps, density plots and time charts.

1. To visualize a summary of the sorting rate of waste through different graphical procedures to provide an overview of the performance of the country.

2. To visualize a summary of the waste production through different graphical procedures to provide an overview of the quantity and composition of it.

This exploration will also inquire on noteworthy regional scenarios, including detailed investigations on the specific condition of the region.

1.2.2 RO2 – Level 2

Techniques of Inferential Statistics such as T-tests and ANOVA will be applied to the dataset to identify relationships and patterns between the waste generation and other potential influencing factors.

1. To assess whether the proportion of the categories of waste produced is the same or not for each region.

2. To seek whether the recycable waste per capita is correlated to its Gross Domestic Product (GDP) per capita.

1.2.3 RO3 – Level 3

Different Machine Learning algorithms will be used to predict the future trends of Italy in waste production and sorting. Of all the available prediction algorithms, two will be compared and evaluated.

1. To analyse the initial distribution to assess which regression algorithm fits the most

2. To predict the kg/per capita of waste produced in the next few years with the proposed regression model

3. To predict the national percentage of sorted waste in next few years with the proposed regression model

1.2.4 RO4 – Level 4

In this section the task of image classification on waste images will be performed. A comparison of the performance of different Deep Learning neural networks will be outlined, discussing loss and accuracy metrics.

1. To solve the task of classification by using the network models DenseNet121 and MobileNet.

2. To improve the accuracy of the task by using the same network models pre-trained on the ImageNet dataset (Transfer Learning).

1.3 Rationale

Italy is the country with the highest percentage of recycling of total waste in Europe. With a rate of 79%, the country recycles more than the double of the EU average, and more that all the other leading European countries such as France (56%), United Kingdom (50%) and Germany (43%) (Symbola 2021).

However, the recycling is only the final phase of a process that starts in everyone's home with waste sorting. Categorizing and classifying the waste that we produce is fundamental to the reuse of such materials in the perspective of a circular economy. Therefore, Italian rates in municipal waste sorting are analyzed to discover trends and factors influencing the waste production, and identify margins of improvement.

1.4 Scoping of Research

The scope of this report is limited to the analysis of the dataset provided by the official government institution ISPRA on waste production in Italy, as well as the analysis and classification of waste images, presented as a solution to the problem of misclassification.

The data cover the production of solid waste in 2020 for 12 different categories. In addition, complementary data will be used to assess the research questions at Section 1.2.2, including data regarding the regional and national Gross Domestic Product (GDP) and data from other EU members to compare the expected standards for Section 1.2.3.

The main focus of this study is to find interesting features and trends of the collected data through EDA, Inferential Statistics, Machine Learning and Deep Learning in terms of regional and national overview of the waste production and sorting of 2020 in the country.

1.5 Contribution to Research

This report aims to provide an overview of solid waste production and sorting in Italy. Through a deep analysis conducted both at national and regional level, the challenges and opportunities of improvements for the country are outlined so that the first step towards a better awareness of the situation is taken. Accounting waste generation rates is fundamental to quantify impacts, plan targeted solutions and set objectives (Wilson & Velis 2015).

One of the latest solutions to automate the process of recycling are Smart Bins, bins enhanced with the ability of real-time classification. Studies have been carried out on Convolutional Neural Network light-weight models to correctly classify images in no more than 6 categories (White et al. 2020), (Chu et al. 2018) and (Ruiz et al. 2019). In this report a dataset of 12 types of waste will be classified through novel light-weight models such as MobileNet, in order to support the basal activity of waste management (Ahmad et al. 2020).

Ultimately, the study conducted aligns with the 17 Sustainable Development Goals (SDGs), established by the 2030 Agenda for Sustainable Development of the United Nations. In particular, it aligns with the twelfth goal: Ensure sustainable consumption and production patterns, and the target five: By 2030, substantially reduce waste generation through prevention, reduction, recycling and reuse (United Nations 2015).

1.6 Organization of the Report

This section provides an overview of the organization of this report. After presenting the problem of solid waste and waste management, and stating the aim and objectives of this study in Chapter 1; an in depth review of current laws on waste management and of the state of the art algorithms for waste segregation is carried out in Chapter 2. Chapter 3 describes the methodology used to perform the analysis on the datasets, which results are discussed in Chapter 4. Finally, evidence-based suggestions are given in Chapter 5 and directives for future developments are proposed in Chapter 6.

Chapter 2

Literature Review

As detailed in (Dieguez 2020), a linear economy is characterized by the three phases of “take-make-dispose”: raw resources are processed and turned into products; products are used until their end-of-life, when they are ultimately discarded as waste.

Linear economy has been the foundation of companies’ business all over the world since the beginning of industrialisation, overshadowed only in the last years by the concurrent concept of circular economy, which aims, on the contrary, to put an end to waste production (Foundation 2015).

As a result, waste has always been the unavoidable consequence of the production processes of the past centuries (Chen et al. 2020) and consumerist behavior of nowadays societies leads to a generation of waste in constant growth (Milonton 2021).

Municipal solid waste is a subdivision of waste, mostly generated by households, whose composition and classification strongly depends on the municipality where it is collected (Nanda & Berruti 2021). According to World Bank, annually, 2.01 billion tonnes of municipal solid waste is produced worldwide, and the value is expected to increase to 2.2 billion tonnes by 2050.

A typical composition of municipal solid waste can be seen in the Table 1, adapted from (Nanda & Berruti 2021).

Every day, an average of 0.74 kg of waste is produced by each one of us, with a range that varies from 0.11 to 4.54 kg (Kaza et al. 2018). In EU countries, citizens produce an average of 1.38 kg per day, as reported by Eurostat.

Waste composition, production, and management has been found positively correlated to income level by multiple studies in literature: (Sivakumar & Sugirtharan 2010), (Magazzino et al. 2021), (Ozcan et al. 2016), (Iyamu et al. 2020), (Mazzanti & Zoboli 2008), (Bandara et al. 2007); that pointed out how richer countries generate more waste, and at the same time recycling more, by producing less organic waste (an average of 32% according to (Kaza et al. 2018)).

Organic waste	Plate waste, spoiled food, vegetable and fruit refuse, kernels, peels, seeds, coffee residues and waste tea leaves
Yard waste	Leaves, grass, trimmings
Paper and cardboard	Newsprint, flyers, magazines, books, tissues, thermal and copy paper, packaging boxes, corrugated fiberboard
Plastic	Single-use plastic containers and utensils, beverages bottles, toys, plastic bags
Metal	Cans, aluminum foil, knives, wires, fences
Glass	Containers, wine and liquor bottles
Electronic waste	Batteries, electronic devices
Miscellaneous	Ceramics, biomedical wastes
Hazardous materials	Pharmaceutical waste, print toners, pesticides, car oil

Table 1: Typical composition of municipal solid waste

Waste management refers to the process of collecting, transporting, recovering (including sorting), and/or disposing waste, where material recovery refers to the preprocessing and preparation of a material in order to be reused, recycled or used as a mean to generate energy (EU 2008). There exist four main clusters of methods of waste disposal, each one including different categories.

1. **Landfill** A landfill is a permanent installation of dumped waste [From: Reference Module in Earth Systems and Environmental Sciences, 2014]. This method is the most popular nowadays, according to Kaza et al. (2018), 70% of waste is dumped. Dumping sites leads to soil and water pollution if the flow of leachate – a solution that generated from fluids present in the waste – is not controlled or prevented (Periathamby 2011).

2. **Incineration/Combustion** This method consists of burning municipal solid waste. As stated by Kaza et al. (2018), the 11% of waste is incinerated.

3. **Recovery and recycling** Recycling is the process of converting by-products into new products to prevent energy usage and consumption of fresh raw materials. The items discarded are processed to extract or recover substances. According to the European Environmental Agency (EEA), the EU recycles an average of 32% of waste, while globally only the 13% is recycled.

4. **Composting** Composting is the biological decomposition of wastes. It is generally achieved by microbiological treatment (Doble & Kumar 2005). Globally, only 5% of the waste is composted Kaza et al. (2018).

In 2020, in the EU, 48% of municipal solid waste was recycled (recycling and composting – recycling of organic matter).

In Italy, for the same year, according to ISPRA (see Section 2.1), 48% of solid waste was recycled (recycling and composting – recycling of organic matter).

2.1 Policies and regulations of solid waste management

European Union

In 2008, the European Union published the Waste Framework Directive, establishing the fundamental notions and definitions of waste management, and setting goals and policies to actively face the problem of waste production and recovery. The core of the EU waste management is a five-step pecking order for handling and disposing waste, the Waste Hierarchy, which can be seen in Figure 1 (EU 2018a).

Figure 1: EU Waste Hierarchy

As stated on the European Commission website;

> Preventing waste is the preferred option, and sending waste to landfill should be the last resort.

To meet the objectives of this Directive, EU nations should take the necessary steps to accomplish the following goal in terms of re-use and recycling of waste materials from households:

- by 2020, the percentage should be increased to a minimum of 50%;
- by 2025, the percentage should be increased to a minimum of 55%;
- by 2035, the percentage should be increased to a minimum of 60%;
- by 2035, the percentage should be increased to a minimum of 55%;

In 2023 the legislation is going to be revised due to the increase in municipal waste generation, especially in waste oils and textile streams. The new revision by the European Commission will focus also on waste grouped collection (European Commission 2008).

Italy

Italy is one of the founders of European Union and therefore it implements EU legislations in national laws as any other EU member.

Italy's political geography is the result of its history and the morphology of its diverse landscape, which also reflects the different cultures and development rates. Every region is unique, yet Italy can be divided into three macro-areas:

- by North Italy we identify regions from North-Ovest and North West, respectively Liguria, Lombardia, Piemonte and Valle d'Aosta, and Emilia-Romagna, Friuli-Venezia Giulia, Trentino-Alto Adige and Veneto;
- by Center or Middle Italy we identify the regions Lazio, Marche, Toscana and Umbria;
- by South Italy we identify regions from South Italy and Insular Italy, respectively Abruzzo, Basilicata, Calabria, Campania, Molise and Puglia, and Sardinia, and Sicily;

Even if some national regulations exists, regions are responsible for defining the metholodogies of waste collection.

This result in many different methods of collections (Section 2.2) that are strictly related to the categories in which waste is sorted.

In Italy the legislation DLGS 152/2006 (Gazzetta Ufficiale 2006) already embraces the waste hierarchy formalized in the Waste Framework Directive two years later. It introduces the Extended Producer Liability, which provides that the costs of managing certain specific waste streams are partially or totally borne by the producers/distributors of the products. The legislation suggests the use of:

1. Economic instruments, eco-balances, environmental certification systems, analysis of the life cycle of products, eco-label system for the assessment of the impact of a specific product on the environment

2. The promotion of agreements and contracts aimed at preventing and reducing the quantity and danger of waste;

3. Initiatives aimed at encouraging the reuse of products and the preparation for the reuse of waste.

The Italian Institute for Environmental Protection and Research, ISPRA (Istituto Superiore per la Protezione e la Ricerca Ambientale), has been instituted by DLGS 112/2008 (Gazzetta Ufficiale 2008). A tool provided by ISPRA is the Waste Register, which, with national, regional and provincial sections, has the responsability of providing a complete and up-to-date framework regarding waste in the country (ISPRA 2021).

The Waste Register (Gazzetta Ufficiale 2006) keeps track of:

1. the quantity of urban waste collected in the territory;

2. the quantity of special waste collected in the territory;

3. the entities managing the waste, specifying the operations carried out, the types and quantity of waste managed by each one;

4. the costs of management and technical and financial depreciation of investments for waste management activities;

5. data relating to separate collection;

6. the quantities collected, broken down by materials.

Professional waste collection and transport activities, recovery and disposal operations must disclose annually the amounts and the details of the trash subjected to the mentioned activities.

With the legislation DLGS 116/2020 more entities producing waste can fall into the category of municipal solid waste (offices, restaurants, museums) if this activities perform waste sorting (Gazzetta Ufficiale 2020).

Furthermore, the use of RENTRI (National Electronic Register for Waste Traceability), already presented in 2019, is better defined.

2.2 Waste segregation

Efficient and precise raw material sorting is the key to the positive outcome of the recycling task (Mao et al. 2021). The waste sorting process begins with the differentiation by the citizens, based on the nature of the single element (Liang & Gu 2021).

Due to the many different material to dispose, it is not always easy to recycle garbage in the correct way. Therefore the need to provide support to this tasks with the use of technology, especially with Computer Vision and Deep Learning techniques, in order to automatically detect and classify waste types (Mao et al. 2021).

The EU directive on Waste Management provides a List of Waste (LoW), regularly revised. It consists of a standardized code to help authorities and countries to correctly manage and categorize waste across Europe. It identifies more than 50 types of waste, between hazardous and non-hazardous waste, and outlines a top-down procedure to identify the right class for the source of disposal (European Commission 2008).

In Italy, the legislation DLGS 116/2020, interpreting the EU regulation, is active from 1/1/2021. It identifies 12 classes of solid waste: organic waste, paper and cardboard, plastic, wood, metal, packaging made of mixed parts, mixed materials, glass, textile, toner, bulky, resins and paints, detergent, other and unsorted waste (Gazzetta Ufficiale 2020).

The municipalities are responsible of defining the individual modalities of conferral, collection and transport of urban waste in order to assure a distinct management of the various waste fractions and promote their recovery. This leads therefore to different implementations of the services that can cause the citizen to misclassify the product. According to a research covering the 90% of the Italian population conducted by Nestlè (Nestlé 2021), depending on the materials, in the country there are from 62 to 93 different solutions for the separate collection of each material. In particular, the different methods concern the various types of containers (bells, sacks, bins), of sizes, as well as the organization of the waste collection service (road, home or ecological island).

As a matter of fact, despite the UNI (Italian National Unification) 11686 standard on "Waste Visual Elements" of 2017, which introduces a way to define municipal waste fractions by means of specific symbols, in some cases different colors are still used for different indications. Indeed, in some cases there are multiple solutions in the same center or municipality (European Standards 2012).

This could constitute an obstacle in the correct differentiation, as industries and municipalities used to decide their own color-coding before the unification.

In Figure 2 is possible to see an example of the situation before the improvement

Figure 2: A comparison of the color scheme for waste collection and sorting before UNI standard

brought by the standard: the variation of the color scheme of two municipalities in region Liguria (the recyling rates of the region will be further detailed in Section 4.1.1) that are only 15km far. In the Figure is also possible to see the color scheme proposed by the standard UNI.

In addition, due to the complexity of waste taxonomies, discerning among municipal solid waste materials has become more challenging for both residents and institutions (Chen et al. 2021).

2.2.1 Smart Solutions

Smart solutions have been implemented over the years to help citizens segregate solid waste. Several applications and web applications that provide information about recycling, collection times, lists of hazardous and non-hazardous materials have been developed to facilitate and support the recycling process (see Table 2).

name	task
iRecycle[1]	recycling ideas
RecycleSMart[2]	council recycling
Recycle![3]	recycling guide and calendar
Junker[4]	recycling guidelines per product

Table 2: Examples of apps available on the market

The use of Artificial Intelligence

Automated Waste Sorting is the process of segregating waste through automation techniques, including Artificial Intelligence and IoT sensors. Computer vision is the most common choice in terms of waste sorting, since it could significantly speed up the manual classification of garbage and reduce the health risks at which sorters are exposed (Gutberlet & Uddin 2017). Increasing the purity of the solid waste is the first step towards a reduced workload in the plants, and to achieve this result it is important to try classify waste in the most correct way the moment it is generated and disposed in houses, offices, shops, and restaurants.

During the last decade different solutions have been proposed, formalized in the concept of the smart bin. Smart bins make use of several technologies such as Radio frequency identification (RFID) readers, ultrasonic and weight sensors, and image recognition algorithms to sort waste (Mukherjee et al. 2021).

The Table 3 summarizes the different Deep Learning and CNN models used to perform garbage classification present in literature so far, since sensors are outside of the scope of this report.

An exhaustive review of all the methods can be found in the study of Lin et al. (2022).

Few implementation and real smart bins exist: research have been focusing more on the evaluation of the performances of the classification algortihm on the different datasets. One example is the smart bin Nando only based on Machine Learning techniques realized by ReLearn[5]. Nando classifies the waste in four categories: glass, paper and cardboard, plastics, and metals. ReSet, instead, produced by the same company, is a tool that installed on the existing bins helps people correctly classifying the solid waste. The tool consists of a camera placed on top of each bin and a dashboard that displays information regarding the piece of waste thrown away, if it is correctly classified or not. Another example of smart bin is Bin-E, a device combining object recognition with IoT sensors to sort and compress the recyclable waste automatically[6].

[1]https://earth911.com/irecycle/
[2]https://www.recyclesmart.com
[3]https://recycleapp.be/
[4]https://www.junkerapp.it/
[5]https://www.re-learn.eu/reset/
[6]https://www.bine.world/customization/

Paper	Model	Task
Rahman et al. (2020)	ResNet34, VGG16, AlexNet, ResNet50	6 classes (cardboard, glass, metal, paper, plastic, trash)
Mao et al. (2021)	DenseNet121 and Genetic Algorithms	6 classes (cardboard, glass, metal, paper, plastic, trash)
Gothai et al. (2022)	12-layer architecture	7 types of plastic (HDPE, PET, LDPE, PVC, PP, PS, other)
Soundarya et al. (2022)	InceptionV3	6 classes (cardboard, glass, metal, paper, plastic, trash)
Sallang et al. (2021)	MobileNetV2	4 classes (glass and plastic, metal, paper and cardboard, non-detectable)
Bobulski & Kubanek (2019)	15-layer architecture (from AlexNet)	7 types of plastic (HDPE, PET, LDPE, PVC, PP, PS, other)
Wang et al. (2020)	ResNet50	6 classes (cardboard, glass, metal, paper, plastic, trash)
Gyawali et al. (2020)	ResNet18	4 classes (glass, paper, plastic, metal)
Srinilta & Kanharattanachai (2019)	VGG16, ResNet50, MobileNetV2, DenseNet121	4 classes (general, compostable, recyclable and hazardous waste)
Alsabei et al. (2021)	ResNet50, VGG16, InceptionV3, Xception	6 classes (cardboard, glass, metal, paper, plastic, trash)

Table 3: List of the use of CNN for waste segregation

Chapter 3

Methodology

The dataset used in this report to describe the Italian scenario in waste generation is the dataset provided by the Italian Institute for Environmental Protection and Research, ISPRA (Istituto Superiore per la Protezione e la Ricerca Ambientale), about the production and recycling of solid waste of 2020 (ISPRA 2021).

3.1 Macro methodology

3.1.1 Data collection

Municipal waste production

The data are collected by ISPRA from the different public entities that are in charge of registering these data at a micro level through a standardized methodology. Nowadays, information is collected from: regions, provinces, regional and provincial agencies for the protection of the environment (ARPA/APPA), and regional and provincial waste observatories.

Missing, partial or incongruent data are integrated through the databases of the MUD (Modello Unico di Dichiarazione ambientale), a communication that organizations and companies must submit annually. Data is collected and elaborated down to the municipal level, unless they are provided per consortium or union of municipalities. A full detailed explanation of the process can be read on the official website of the institute.

Garbage images

The dataset is a public dataset available on Kaggle under the name of *Garbage Classification (12 classes)*. The images are collected by combining other public dataset available on the platform and web scraping. Further informations are availble on the website.

3.1.2 Data dictionary

Municipal waste production

The dataset *Production and recycling of municipal scale, year X (ISPRA)* consists of yearly overviews of the solid waste produced by the villages in Italy from 2012 to 2020. Each dataset is made of 25 columns and each row is an observation of the municipal waste production, specified by the following attributes:

1. IstatComune: a unique ID for each city hall
2. Regione: nominal variable, specifies the region of each village
3. Provincia: nominal variable, specifies the province of each village
4. Comune: nominal variable, specifies the name of each village
5. Popolazione: continuous variables, specifies the inhabitants of each village
6. Dato riferito a: nominal variable, specifies the typology of municipality (village or union of villages)
7. Frazione umida (t): continuous variables, specifies the amount of organic fraction produced by each village in tons
8. Verde (t): continuous variables, specifies the amount of organic fraction produced by each village in tons
9. Carta e cartone (t): continuous variables, specifies the amount of paper and cardboard waste produced by each village in tons
10. Vetro (t): continuous variables, specifies the amount of glass waste produced by each village in tons
11. Legno (t): continuous variables, specifies the amount of wood waste produced by each village in tons

12. Metallo (t): continuous variables, specifies the amount of metal waste produced by each village in tons

13. Plastica (t): continuous variables, specifies the amount of plastic waste produced by each village in tons

14. RAEE (t): continuous variables, specifies the amount of WEEE produced by each village in tons

15. Tessili (t): continuous variables, specifies the amount of textiles produced by each village in tons

16. Selettiva (t): continuous variables, specifies the amount of selective waste produced by each village in tons (pharmaceuticals, T/FC containers, batteries and accumulators, paints, inks and adhesives, vegetable oils and mineral oils)

17. Rifiuti da C e D (t): continuous variables, specifies the amount of construction and demolition waste produced by each village in tons

18. Pulizia stradale a recupero (t): continuous variables, specifies the amount of waste collected from cleaning the streets produced by each village in tons

19. Ingombranti misti a recupero (t): continuous variables, specifies the amount of recycling bulky waste produced by each village in tons

20. Altro (t): continuous variables, specifies the amount of non-classified waste produced by each village in tons

21. Totale RD (t): continuous variables, specifies the total amount of recycling waste produced by each village in tons

22. Ingombranti a smaltimento (t): continuous variables, specifies the amount of non-recycling bulky waste produced by each village in tons

23. Indifferenziato (t): continuous variables, specifies the amount of undifferentiated waste produced by each village in tons

24. Totale RU (t): continuous variables, specifies the total amount of solid municipal waste produced by each village in tons

25. Percentuale RD (%): continuous variables, specifies the percentage of recycling of the total waste produced by each village

Garbage images

The dataset contains 15,150 images from 12 different classes of household waste: paper, cardboard, biological, metal, plastic, green-glass, brown-glass, white-glass, clothes, shoes, batteries, trash.

3.1.3 Data pre-processing

Municipal waste production

The dataset presents missing values in the columns that refer to the amount of waste produced. In Figure 3 is possible to see the percentage of missing data for each column of the dataset.

As will be discussed more in Section 4.2, the amount of waste produced is positively correlated with the number of inhabitants, therefore traditional imputation methods such as KNN algorithms or replacing the values with the mean/median/mode, could not be used to fill the missing values of the dataset.

```
IstatComune                          0.000000
Regione                              0.000000
Provincia                            0.000000
Comune                               0.000000
Popolazione                          0.000000
Dato riferito a                      0.000000
Frazione umida(1) (t)               12.817917
Verde (t)                           26.686069
 Carta e cartone (t)                 3.112742
Vetro (t)                            4.213590
Legno (t)                           28.799190
Metallo (t)                          8.743515
Plastica (t)                         3.935214
RAEE (t)                            11.299507
Tessili (t)                         28.849804
Selettiva (t)                       12.615462
Rifiuti da C e D (t)                41.933443
Pulizia stradale a recupero (t)     59.002910
Ingombranti misti a recupero (t)    13.728964
Altro (t)                           22.143490
Totale RD (t)                        2.176389
Ingombranti a smaltimento (t)       77.831203
Indifferenziato (t)                  1.923320
Totale RU (t)                        1.910667
Percentuale RD (%)                   2.176389
dtype: float64
```

Figure 3: Dataset missing values

The following procedure has been used to replace missing values: the values of each municipality are replaced with the values of the municipality of the same region or province with the closest number of inhabitants, in proportion per number of inhabitants. This method is used to fill the missing values from columns of waste production, while the totals and the percentage are obtained by summation and proportion respectively. The reason why this method has been adopted is because each province/region has different regulations and procedures in terms of recycling.

Note that values regarding municipalities which are denoted to refer to an aggregation of municipalities are not filled, as they are meant to miss on purpose, in order to appear only once in the dataset.

Garbage images

The dataset is splitted into training set, validation set, and test set as follows: 80% train set, 10% validation set, and 10% test set.

3.2 Micro methodology

The analysis of the designated dataset is carried out in three different layers of depth.

The first layer makes use of Descriptive Statistics to describe the dataset. Mean, median, mode, standard deviation, and upper-quartile are used to describe the features of the dataset and answer the the research questions detailed in Section 1.2.1. This analysis in- cludes multiple visualizations such as box plots, bar charts, histograms, maps, tree maps, and density plots.

The second layer makes use of Inferential Statistics to describe the dataset. T-test and ANOVA are used to answer the research questions detailed in Section 1.2.2.

The third layer makes use of Machine Learning algorithms to predict features of the dataset. Polynomial Regression and Random Forest are applied to predict the features outlined in Section 1.2.3.

Finally, the image dataset is taken into account to perform the last analysis detailed in Section 1.2.4.

The fourth layer makes use of Deep Learning algorithms to classify garbage images. Two different Convolutional Neural Networks are applied to the dataset and their performances are compared through Accuracy and Loss metrics. Transfer learning techniques are also used to improve the network performance.

MobileNet

MobileNets is a family of CNN developed specifically for mobile and embedded vision applications. Therefore, this family of models is suitable for the implementation of intelligent waste management system. The strenght of this architecture is high efficiency obtained with a reduced computational cost Howard et al. (2017). The structure of MobileNet (v1) can be seen in Figure 4

Type / Stride	Filter Shape	Input Size
Conv / s2	$3 \times 3 \times 3 \times 32$	$224 \times 224 \times 3$
Conv dw / s1	$3 \times 3 \times 32$ dw	$112 \times 112 \times 32$
Conv / s1	$1 \times 1 \times 32 \times 64$	$112 \times 112 \times 32$
Conv dw / s2	$3 \times 3 \times 64$ dw	$112 \times 112 \times 64$
Conv / s1	$1 \times 1 \times 64 \times 128$	$56 \times 56 \times 64$
Conv dw / s1	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 128$	$56 \times 56 \times 128$
Conv dw / s2	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 256$	$28 \times 28 \times 128$
Conv dw / s1	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 256$	$28 \times 28 \times 256$
Conv dw / s2	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 512$	$14 \times 14 \times 256$
5× Conv dw / s1 Conv / s1	$3 \times 3 \times 512$ dw $1 \times 1 \times 512 \times 512$	$14 \times 14 \times 512$ $14 \times 14 \times 512$
Conv dw / s2	$3 \times 3 \times 512$ dw	$14 \times 14 \times 512$
Conv / s1	$1 \times 1 \times 512 \times 1024$	$7 \times 7 \times 512$
Conv dw / s2	$3 \times 3 \times 1024$ dw	$7 \times 7 \times 1024$
Conv / s1	$1 \times 1 \times 1024 \times 1024$	$7 \times 7 \times 1024$
Avg Pool / s1	Pool 7×7	$7 \times 7 \times 1024$
FC / s1	1024×1000	$1 \times 1 \times 1024$
Softmax / s1	Classifier	$1 \times 1 \times 1000$

Figure 4: Structure of MobileNet (v1

This model would be suitable for embedded devices such as Smart Bins thanks to its light weight of only 16MB and its performances of accuracy (top-5) on the ImageNet validation dataset of 89.5% Chollet et al. (2015).

DenseNet121

DenseNets is a family of feed-forward CNN models, where each layer is fully connected to the next one, to make sure the information flow between layers in the network is at its maximum. Specifically, DenseNet121 is a network made of four dense blocks, whose structure is specified in Figure 5 Huang et al. (2017).

Layers	Output Size	DenseNet-121
Convolution	112×112	
Pooling	56×56	
Dense Block (1)	56×56	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$
Transition Layer (1)	56×56	
	28×28	
Dense Block (2)	28×28	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
Transition Layer (2)	28×28	
	14×14	
Dense Block (3)	14×14	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$
Transition Layer (3)	14×14	
	7×7	
Dense Block (4)	7×7	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 16$
Classification Layer	1×1	

Figure 5: Structure of DenseNet121

This model would be suitable for embedded devices such as Smart Bins thanks to its light weight of only 33MB and its high performances of accuracy (top-5) on the ImageNet validation dataset of 92.3% Chollet et al. (2015).

Transfer Learning

Transfer learning consists of in trasfering the knowledge one model adquired from a dataset, and implanting it on different model to solve a similar problem Abadi et al. (2015). The transfer learning procedure is the following:

1. Add the layer from a previously trained model to the new model
2. Freeze the layer so that the knowledge is not forgotten during the training phase of the new model
3. Train the layers on top of the frozen layers, so that this layer can turn the old knowledge into new features for the new problem to solve

Chapter 4

Results and Discussion

After the pre-processing phase, the dataset used to perform the analysis defined in Section 1.2 is shown in Figure 6.

Regione	Provincia	Comune	Popolazione	Dato riferito a	Frazione umida(1) (t)	Verde (t)	Carta e cartone (t)	...	Selettiva (t)	Rifiuti da C e D (t)	Pulizia stradale a recupero (t)	Ingombranti misti a recupero (t)	Altro (t)	Totale RD (t)	Ingombranti a smaltimento (t)
Piemonte	Torino	AGLIE'	2548	Comune	273.774	396.303	92.426	...	1.185	3.889	18.460000	45.331	3.157	1042.158	48.506589
Piemonte	Torino	AIRASCA	3569	Comune	177.232	144.097	273.497	...	3.209	14.730	9.972936	92.150	0.025	1253.957	0.097767
Piemonte	Torino	ALA DI STURA	448	Comune	2.423	9.901	32.738	...	0.127	0.139	9.348720	34.229	0.003	142.306	15.106281
Piemonte	Torino	ALBIANO D'IVREA	1650	Comune	165.909	20.679	70.587	...	0.804	21.220	2.080000	14.499	3.040	480.687	0.052693
Piemonte	Torino	ALMESE	6448	Comune	622.074	1105.521	366.851	...	8.455	74.185	49.396000	118.138	5.475	3054.947	59.353098

Figure 6: Head of the dataset used for the analysis

4.1 EDA: first level of analysis

4.1.1 Recycling rates

In 2020, Italy recycled the 63% of the municipal solid waste produced, as shown in Figure 7, ovetaking by 13% the objective for solid waste recycling set by the EU for the same year and getting closer to the objective of 65% set for 2035 (EU 2008, 2018a).

Both for historical and morphological reasons, it is relevant to look at the rates of the three macro-areas of Italy mentioned in Section 2. This three macro-region contribute to the national reuse of solid waste with different weights, as shown in the bar plot in Figure 7. The North is the macro-region that recycles the highest amount of the solid waste produced (71%), followed by Center (60%) and South Italy (54%).

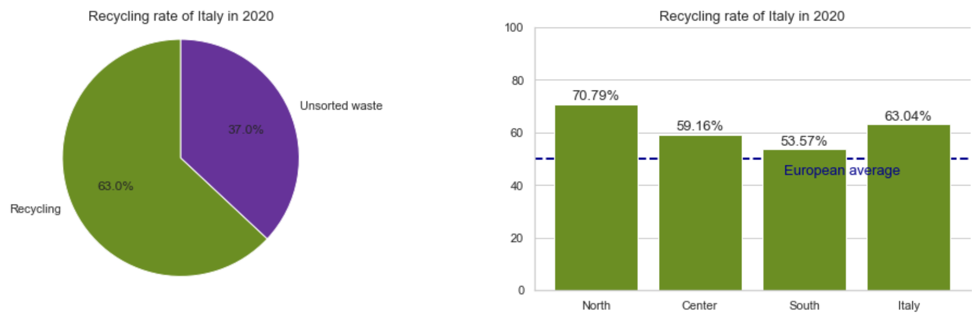

Figure 7: Percentage of solid waste recycled in Italy in 2020

Furthermore, if we look at the regional percentages of recycling, it is even more evident how heterogeneous the Italian scenario is.

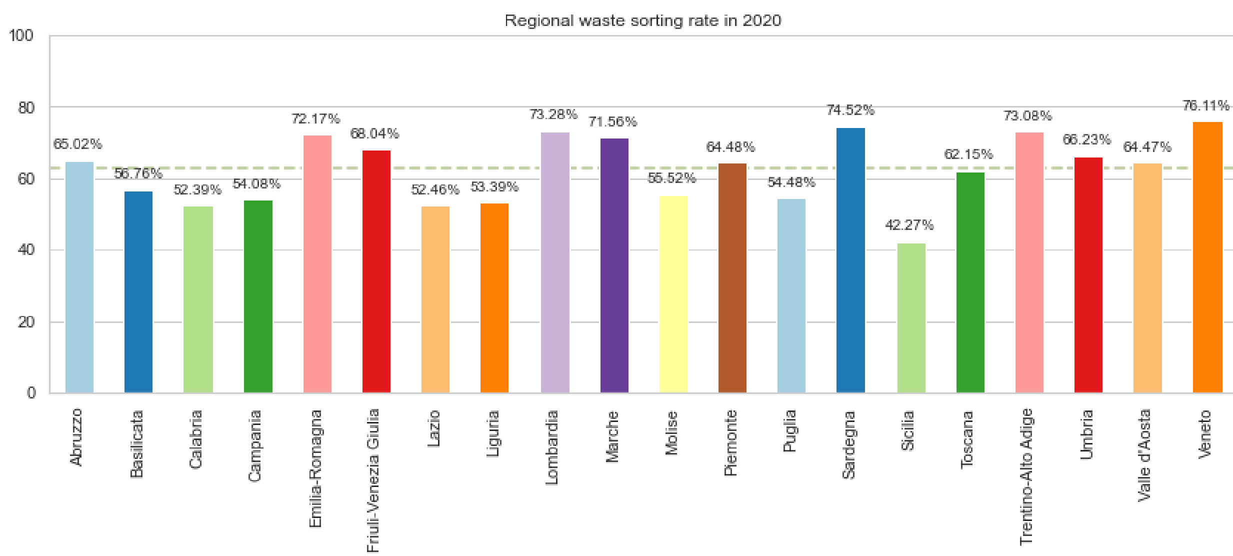

Figure 8: Percentages of solid waste recycled in regions of Italy in 2020

Sicily is the region with the lowest recycling rate (42%), while Veneto is the region that recycles the most: almost the 80%. Five regions out of twenty recycle are in the 75° percentile, recycling more than 71.7% of their municipal waste: Lombardy, Emilia-Romagna, Sardinia, Veneto and Trentino Alto Adige.

From the analysis of the individual region's distribution of recycling rate, Liguria, the northern region that has the lower recycling rate (53%), showed a slighlty bimodal distribution, as seen in Figure 9.

Most of the municipalities in this region have a high recycling rate, but a relevant

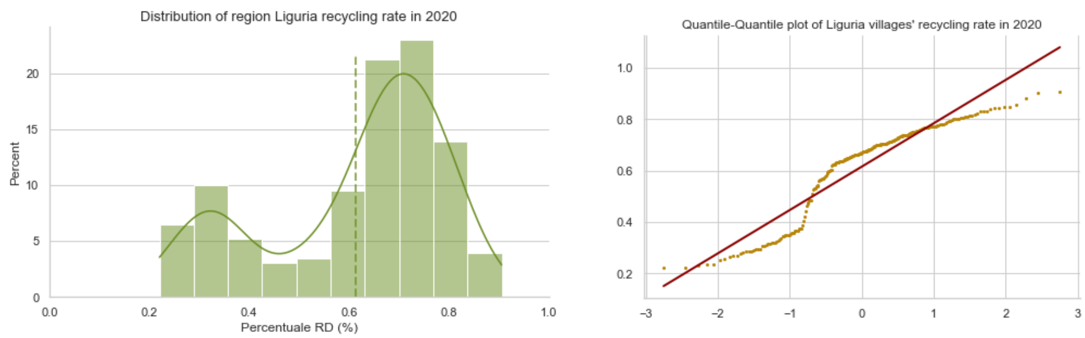

Figure 9: Distribution of municipalities' recycling rate in Liguria

amount of cities (about 30%) has a lower recycling rate. The reason behind this behavior could be linked to the provinces of Liguria, whose recycling rates are shown in Figure 10.

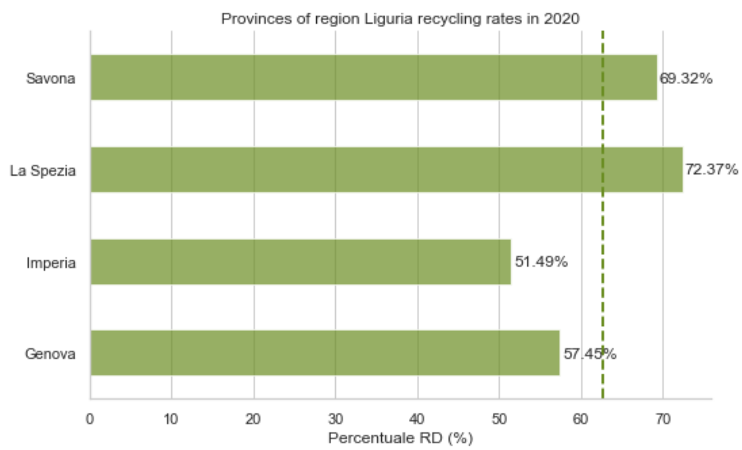

Figure 10: Percentages of solid waste recycled in the provinces of Liguria

Imperia and Genova are the provinces that recycle the least. From the map shown in Figure 11 is clear how the two provinces both have a clusters (the two yellow small groups) of villages that significantly have lower recycling rates.

The reason behind this data is in the lack of proper facilities for the closure of the waste cycle, such as a plant for the mechanical biological treatment of waste or a plant for the treatment of the wet fraction. Therefore, Genoa's province is till sending a large part of the waste produced by citizens outside the Region (Legambiente & Liguria 2021).

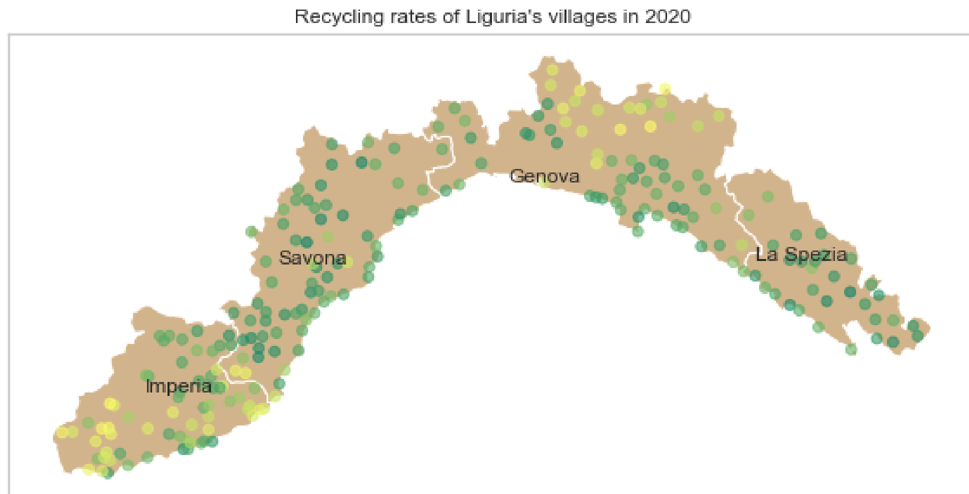

Figure 11: Map of the recycling rates of the municipalities of Liguria

4.1.2 Waste production

Although the high recycling rate of a macro-region or region is a good starting point, it is also fundamental to investigate the amount of municipal solid waste produced. Overall, in 2020, Italy generated 28,9 millions tonnes of municipal solid waste, resulting in circa 488kg of waste per capita.

While the recycling rate varies significantly between the macro-regions, in Figure 12 is possible to see how the production of solid waste per capita only varies of a maximum of +-50kg between North, Center and South Italy.

Further detailing the analysis, in Figure 13 the kg of waste per capita produced by each region of Italy both for recycling and non-recycling material can be observed and compared to the European average for the same year (505kg).

It is interesting to identify the regions that constitute a good trade-off between recycling rate and waste per capita. This is obtained by considering the regions that produce less non-recycable waste, as shown in Figure 14. Regions such as Lombardia, Veneto, ad Sardinia which, as mentioned in 4.1.1 are in the 75° percentile for the recycling rate, are a good tradeoff and represent the tendency toward which other regions should inspire to reach similar values.

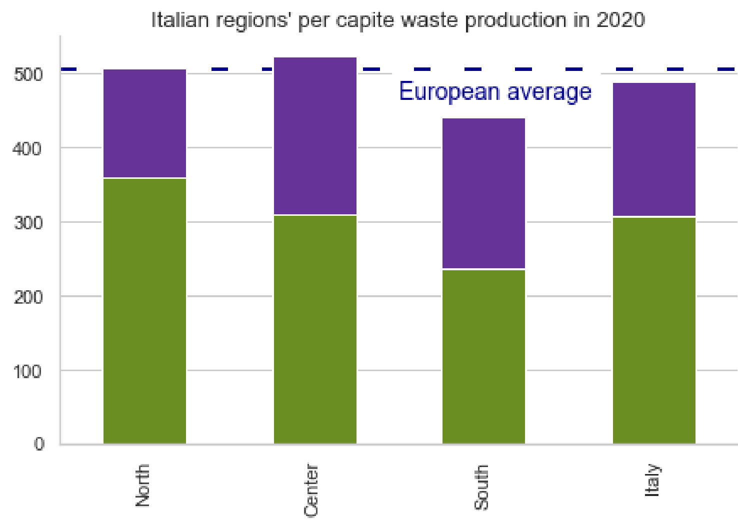

Figure 12: National recycable and unsorted solid waste per capita in kg

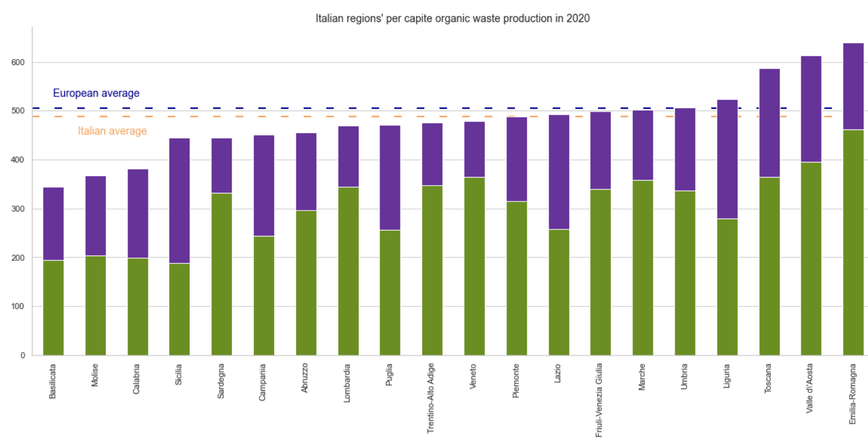

Figure 13: Regional recycable and unsorted solid waste per capita in kg

4.1.3 Waste composition

From the analysis of the dataset referring to the year 2020, the solid waste is mostly made of humid organic waste (39%), followed by paper and cardboard (19%) and glass (12%). The fractions are percentages against the total waste produced in Italy. This overview at national level can be confronted with the regional level, to see which regions contribute the most with which waste category.

From Figure 16, is clear how some materials are produced more in certain areas of Italy. While the rank is maintained for all the three macro-areas, we can notice

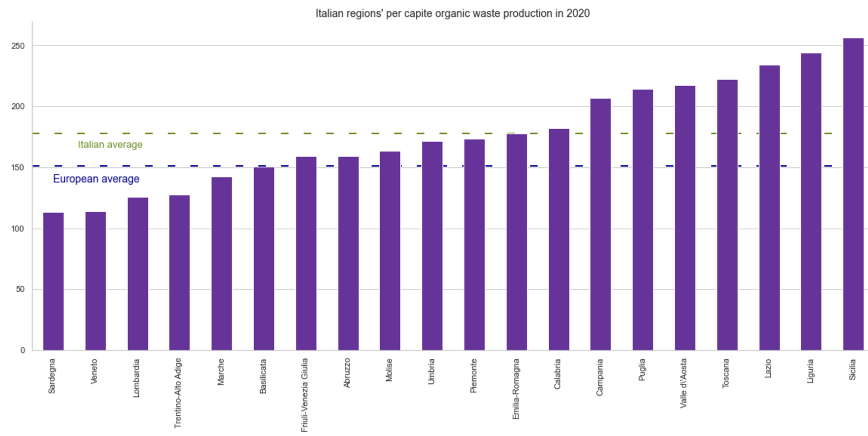

Figure 14: Macro-regional unsorted solid waste per capita in kg

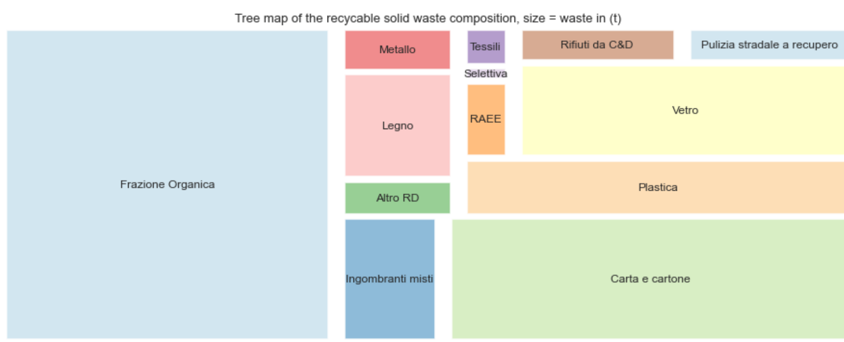

Figure 15: Recycable solid waste composition

how glass is produced at the same percentages, but when it comes for the the humid organic waste the south regions produce more, and when we talk about paper and cardboard the region in the center of Italy consume more.

4.2 Inferential Statistic: second level of analysis

4.2.1 Solid waste composition

As mentioned in Section 4.1.3, the objective is to assess if the composition of solid waste produced in each region varies significantly or not.

In Figure 16 is possible to see the distribution of the percentage composition of four of the categories of the dataset per macro-region[1]. For some of the mentioned

[1]The percentage is the result of the total amount of category-specific waste produced over the total

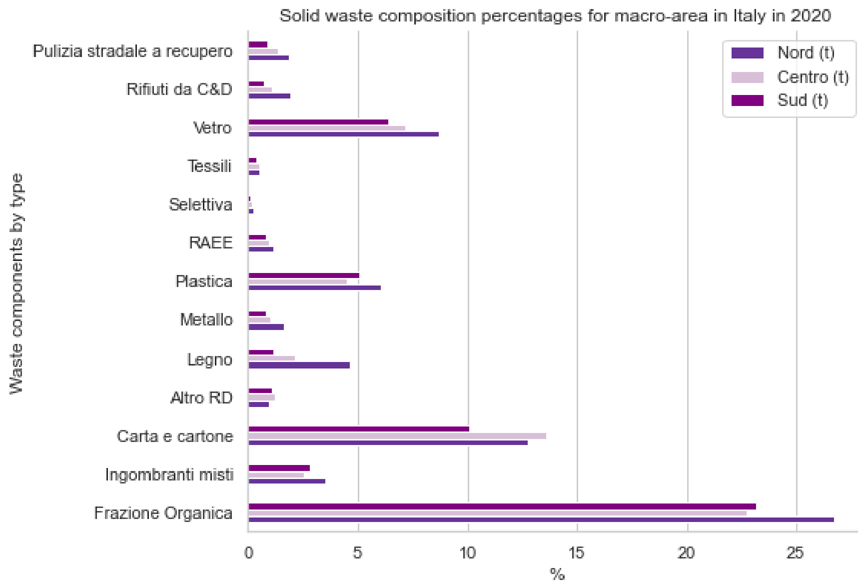

Figure 16: Distribution of waste composition over macro-regions

materials there are some significant differences at a macro-regional level, such as for wood and for organic fraction.

To assess formally what can be read from the histogram, an ANOVA test is carried out on the percentages[2] of composition of the categories paper and cardboard, wood, glass and organic waste considering the percentages of the single regions of the macro-area.

To compare the percentages through a one-way ANOVA we must define the hypothesis as follows:

> H_0: $\mu_1 = \mu_2 = \mu_2$ (the mean of the percentages of waste category produced in each macro-region is equal)
>
> H_A: at least one $\mu_i \neq \mu_j$ (the mean of the percentages of waste category produced is not equal for at least one macro-region)

The confidence interval chosen for the test is 95%, consequently the test is performed with the α-value equal to 0.05.

The output of the tests performed on the dataset can be seen in Figure 17, 18, 19, and 20.

amount of waste produced in that macro-area

[2]The reason why we consider the percentage and not the kg per capita is to avoid the bias of amount of waste produced - who produces more, will have of course a higher production also in a specific category.

```
ANOVA test on Wood
p value = 1.13663e-05
f value = 23.9422
The null hypothesis can be rejected
```

Figure 17: Result of the ANOVA test performed on wood waste

```
ANOVA test on Glass
Variance ratio: 5.511470101526159
```

Figure 18: Result of the ANOVA test performed on glass waste

```
ANOVA test on Paper
p value = 0.6843
f value = 0.387952
The null hypothesis is accepted
```

Figure 19: Result of the ANOVA test performed on paper waste

```
ANOVA test on Organic Waste
p value = 7.68972e-07
f value = 36.0373
The null hypothesis can be rejected
```

Figure 20: Result of the ANOVA test performed on organic waste

The assumption with ANOVA test is that the Variance Ratio should be less than 4 to perform the analysis. For this reason no test was conducted on the glass waste. The other tests show how there is no significant difference in the production of paper, while in the production of wood and organic waste one macro-region resulted to produce more (or less) amount.

Taking into account, for example, the wood waste percentages of the regions, seen in Figure 21, it is clear how the northern regions (in white) produce more wood waste than the rest of Italy, especially South Italy (in red). This is due to the presence of the Alps, mountains range in North Italy, and the wide wooded area.

The region that stands out is the region of Valle D'Aosta, being its wood waste the 8% of its municipal waste production. In Valle d'Aosta, forests occupy 30% of the territory and a great artisan tradition was developed during the centuries.

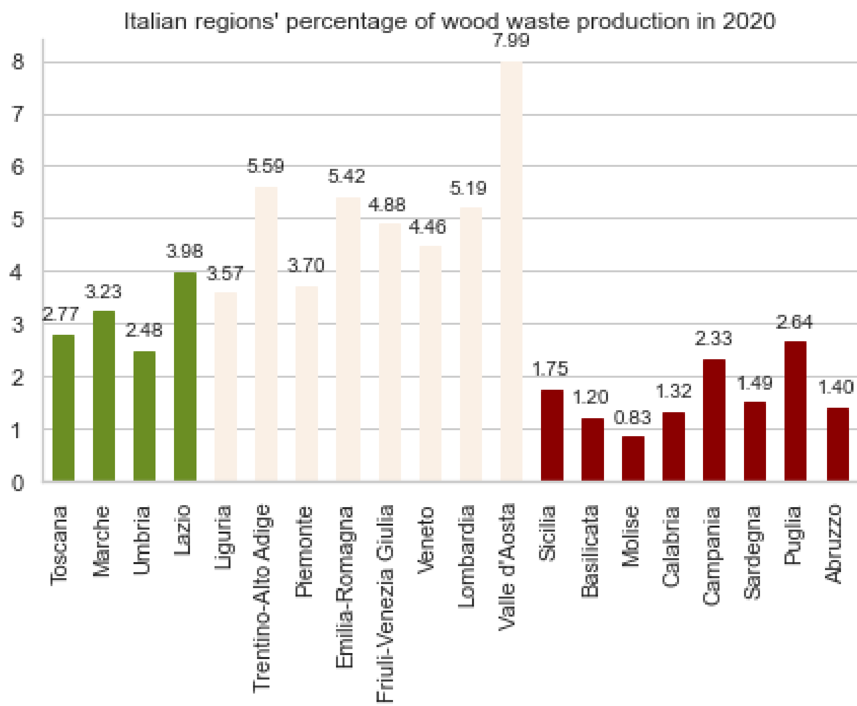

Figure 21: Regional percentages of wood waste production (over the total amount of waste)

4.2.2 Correlation

To better understand the dataset and discover relationship between variables, a correlation matrix is presented in Figure 22

From the Figure 22 is clear that the amount of waste recycled pro capite has a positive correlation with the Gross Domestic Product (GDP), since the correlation value is equal to 0.72. The positive linear correlation is justified by the tendency of richer regions to invest more money in innovative technologies to recover waste. An overview of the GDP per region is shown in Figure 23. Northern regions have higher GDP, while southern region have a lower GDP. According the report GreenItaly, elaborated by da Fondazione Unioncamere and Fondazione Symbola (Symbola 2021), more than the 45% of the companies of North Italy investmented in green processes and products between 2015 and 2019. In the Center of Italy the value reached 20% and in South Italy 29%.

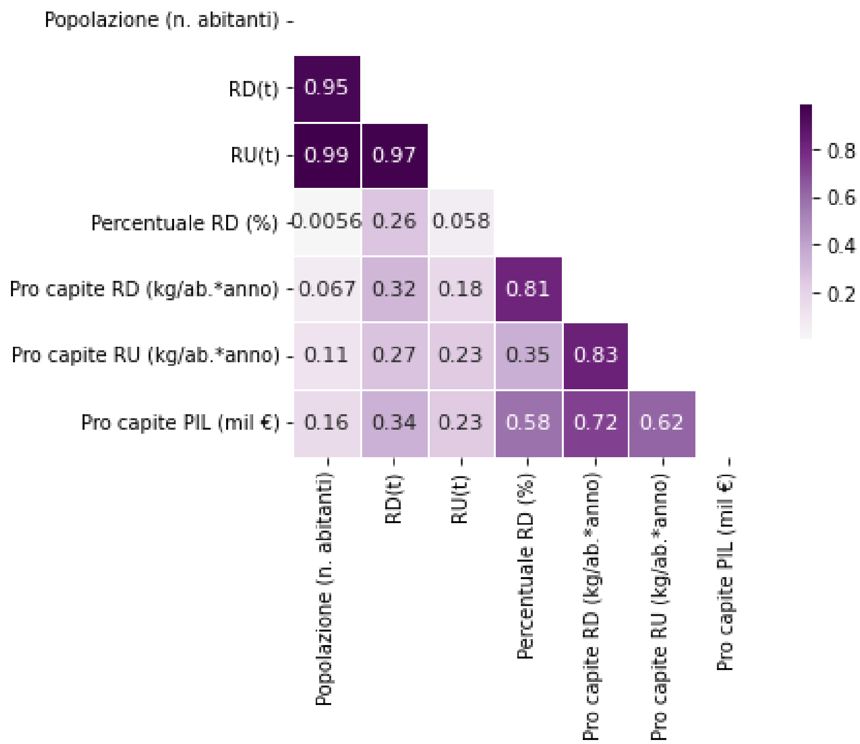

Figure 22: Heatmap of correlation among attributes

4.3 Machine Learning: third level of analysis

4.3.1 Waste production

As mentioned in Chapter 1, the objective is to predict the amount of kg/per capita of waste produced by Italian citizen.

In Figure 24 is possible to see the distribution of the data from 1996 to 2020 regarding the waste production of kg/per capita. The highest peak of 556 kg/per capita is reached in 2006 and then a rapid decrease is registered. This is most certainly due to the publication of the EU law regarding waste management, followed by the interpretation of the Italian Government published in the same year (Gazzetta Ufficiale 2006).

According to the distribution of the data, a Polynomial Regression is applied to perform predictions on the data for the years 2021, 2022, 2023, 2024, until 2025. A Polynomial regression with 2 degrees of freedom (quadratic) is used. The Figure 25 shows that the R-squared error of this model is 0.77 or 77% which can be considered

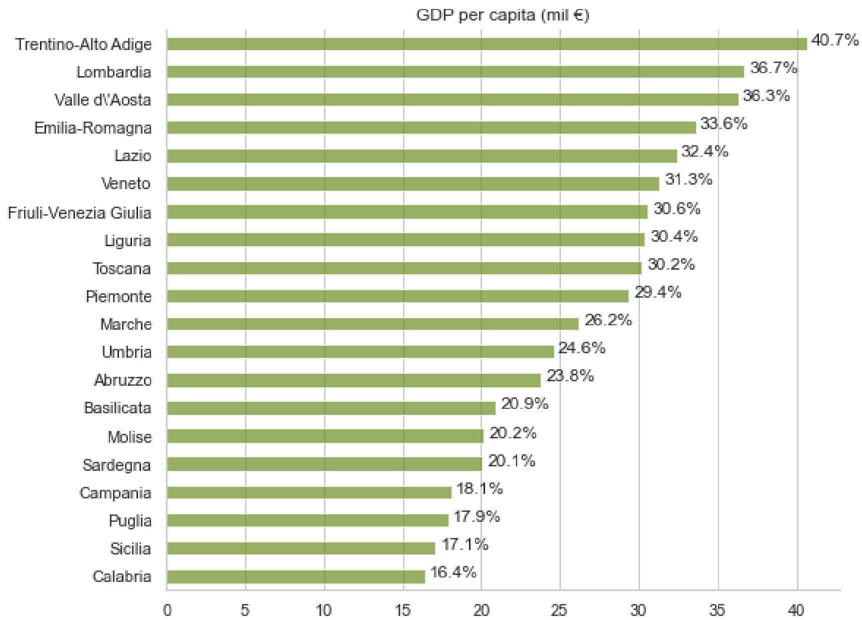

Figure 23: Regional GDP of Italy in 2020

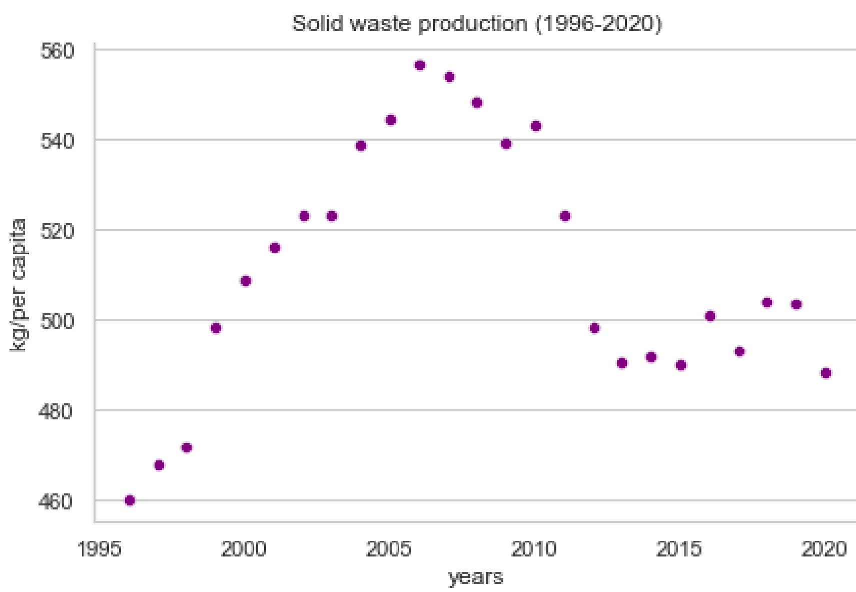

Figure 24: National waste production per capita trend

a good fit. The adjusted R-squared value is also 0.865 which means the attributes chosen to predict are a very good choice and perfectly fitted with the model.

Therefore, the formula to predict the waste production per capita is the following:

$$y = 1.72 * 10^3 x - 0.43x^2$$

```
Model coefficients are: [ 0.00000000e+00  1.72746925e+03 -4.30130001e-01]
Mean Absolute Error, MAE is  11.093402440812543
Mean Squared Error, MSE is  172.30123719750995
Root Mean Square Error is  13.126356585035694
Coefficient of determination or R squared value is  0.7757097691477188
```

Figure 25: Model coefficients and errors

The same behavior is analyzed per macro-region and shown in Figure 26. The Center Italy is the macro-region that, over the years, has been producing more waste (see Figure 12 in Section 4.1.2 for 2020 data) and it is also expected to decrease at a higher rate. Compared to the production of Center Italy, the one of North Italy does not vary as much, in a range of 464-551 kg/per capita. South Italy is the only macro-area that did not show a positive trend for the years 2015-2019. The coefficients and the R-squared (>=0.72) values of the models are shown in Figure 27, 28, 29 respectively, following the equation

$$y = \alpha + \beta x + \gamma x^2$$

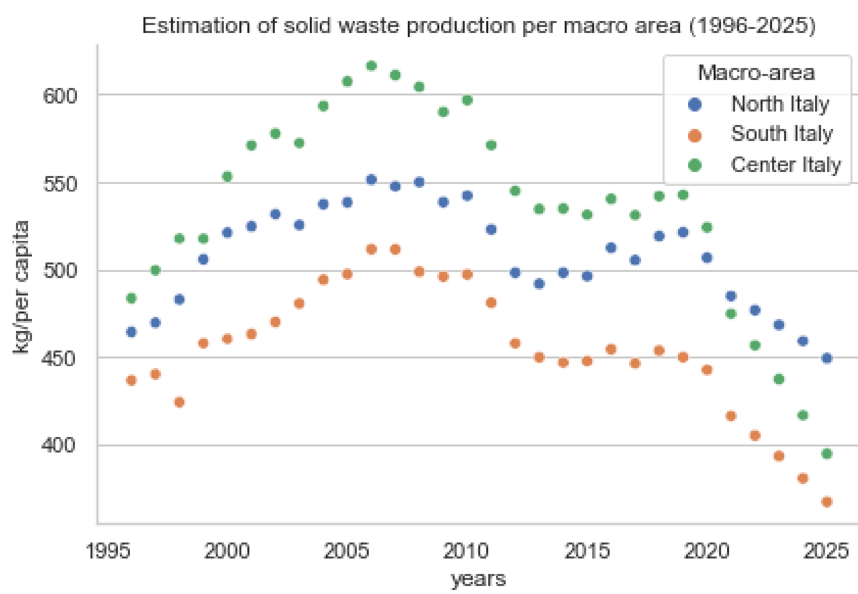

Figure 26: Macro-regional waste production per capita future trend

```
Model coefficients are: [ 0.00000000e+00  1.23456731e+03 -3.07325370e-01]
Mean Absolute Error, MAE is  12.524061765277565
Mean Squared Error, MSE is  202.55834980359458
Root Mean Square Error is  14.23229952620428
Coefficient of determination or R squared value is  0.7201189972289177
```

Figure 27: Model coefficients and errors for North Italy

```
Model coefficients are: [ 0.00000000e+00  2.64861609e+03 -6.59573085e-01]
Mean Absolute Error, MAE is  17.7003650918635
Mean Squared Error, MSE is  397.0559153399963
Root Mean Square Error is  19.926261951003163
Coefficient of determination or R squared value is  0.7586789571609389
```

Figure 28: Model coefficients and errors for Center Italy

```
Model coefficients are: [ 0.00000000e+00  1.61526386e+03 -4.02262403e-01]
Mean Absolute Error, MAE is  7.552503971910687
Mean Squared Error, MSE is  133.5068787821111
Root Mean Square Error is  11.55451767847153
Coefficient of determination or R squared value is  0.7735984001450235
```

Figure 29: Model coefficients and errors for South Italy

The same procedure has been applied to regional level considering two examples: Liguria, whose situation was already presented in Section 4.1.1, and Molise, which is one of the regions that produces less amount of waste per capita, as shown previously in Figure 14.

For what concerns Region Liguria's prediction resulted more accurate with a Polynomial Regression with 3 degrees of freedom, following the equation:

$$y = \alpha + \beta x + \gamma x^2 + \delta x^3$$

The model's values are shown in Figure 30. The R-squared value is equal to 0.92, which denotes a high accuracy in the prediction. The prediction is shown in Figure 31

```
Liguria
Model coefficients are: [ 0.00000000e+00  4.18957468e+05 -2.07945202e+02  3.44031088e-02]
Mean Absolute Error, MAE is  8.344399976651834
Mean Squared Error, MSE is  122.90963902917251
Root Mean Square Error is  11.086461970762922
Coefficient of determination or R squared value is  0.9204454057828358
```

Figure 30: Model coefficients and errors for Region Liguria

For what concerns Region Molise's prediction resulted more accurate with a Polynomial Regression with 2 degrees of freedom, following the formula denoted above.

The model's values are shown in Figure 32. The R-squared value is 0.75, which denotes a good accuracy in the prediction. The prediction is shown in Figure 33.

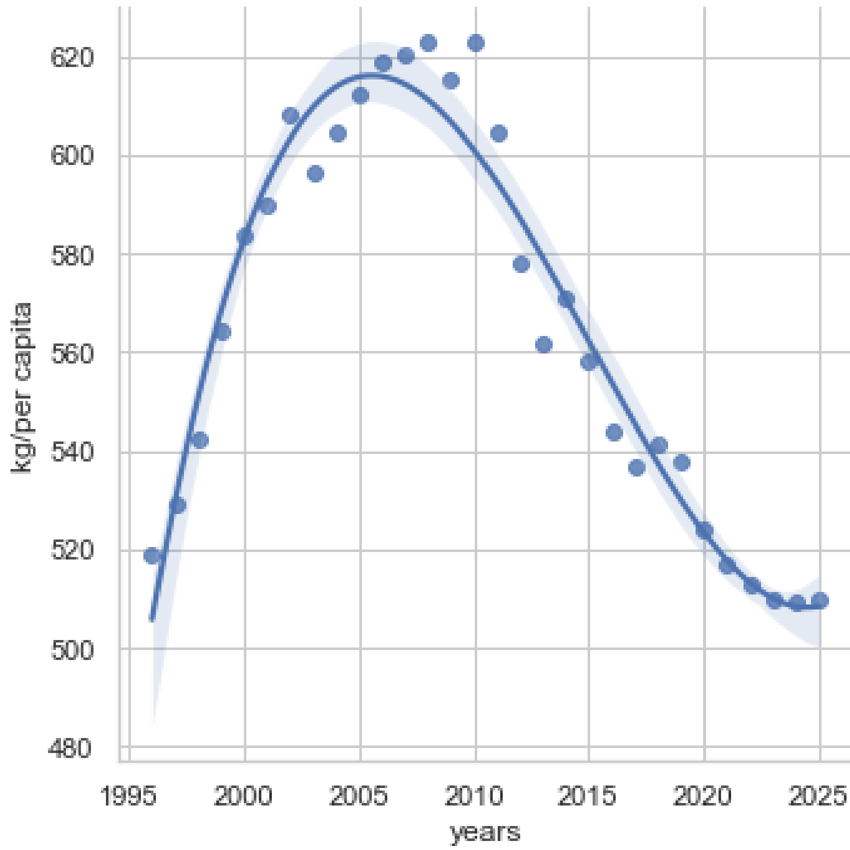

Figure 31: Waste production per capita trend for Region Liguria

```
Molise
Model coefficients are: [ 0.00000000e+00  1.37879161e+03 -3.43158520e-01]
Mean Absolute Error, MAE is  14.64131800985732
Mean Squared Error, MSE is  272.58636481285146
Root Mean Square Error is  16.510189726736986
Coefficient of determination or R squared value is  0.7576876546535846
```

Figure 32: Model coefficients and errors for Region Molise

Both the models are expected to reach a production equal or lower to the one of 25 years ago (1995).

4.3.2 Recycling rates

As mentioned in Chapter 1, the objective is to predict the national percentage of sorted waste in the next decade

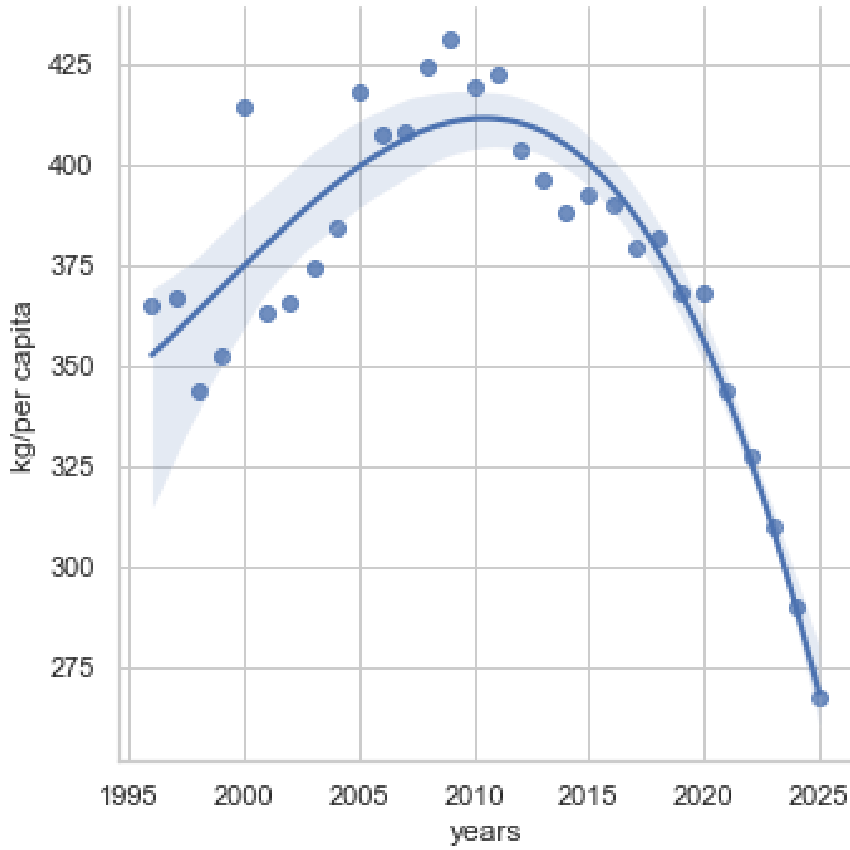

Figure 33: Waste production per capita trend for Region Molise

In Figure 34 is possible to see the evolution of the recycling rate of solid waste from 1996 to 2020. The graph shows a constant, linear behavior. Therefore, it is possible to predict future recycling rates with a Linear Regression model.

The metrics of the model can be seen in Figure 35: the R-squared is equal to 0.92 or 92%, which means that the model fits the data in a precisely. The formula of linear regression is

$$y = \alpha + \beta x$$

.

Thanks to this model, predictions can be made to estimate the recycling rate up to 2025. In Figure 36 it is possible to see the predictions of the model for those years.

The same trend is analyzed per macro-region and shown in Figure 37.

North, Center and South Italy have been predicted through Ploynomial Regression with 2 degrees of freedom, to achieve a higher R-square value as shown in Figure 38,

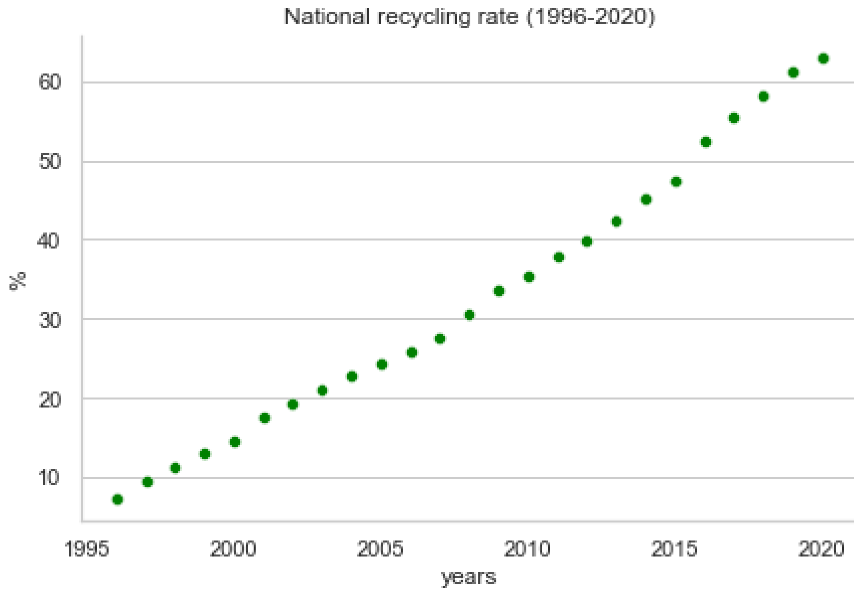

Figure 34: National waste production per capita trend

```
Model coefficients are: [2.32225234]
Mean Absolute Error, MAE is  1.0500550202070684
Mean Squared Error, MSE is  1.7843619103403825
Root Mean Square Error is  1.3358001011904372
Coefficient of determination or R squared value is  0.9254197910386048
```

Figure 35: National waste production per capita trend

39 and 40.

The North Italy is the macro-region that, since the availability of the data, has always been recycling at a higher percentage. Nevertheless the predictions show how the potential performances of South regions, that have been recycling at an exponential rate, could reach higher values in the near future.

Finally, some regional scenarios are analysed: Veneto, as it is one of the regions that recycles the most, while the data referring to Sicily are presented, as it is the region with the lowest recycling rate (as seen in Section 4.1.1).

For what concerns Region Veneto, a Quadratic Polynomial Regression model has been applied, generating the metrics shown in Figure **??** and the estimations in Figure 42. It is possible to see how the rate seems to tend to a limit of 80%, which could be the limit at which is possible to recycle nowadays with the current technologies.

For what concerns Sicily, a Cubic Polynomial Regression model has been applied, generating the metrics shown in Figure **??** and the estimations in Figure 44. It is

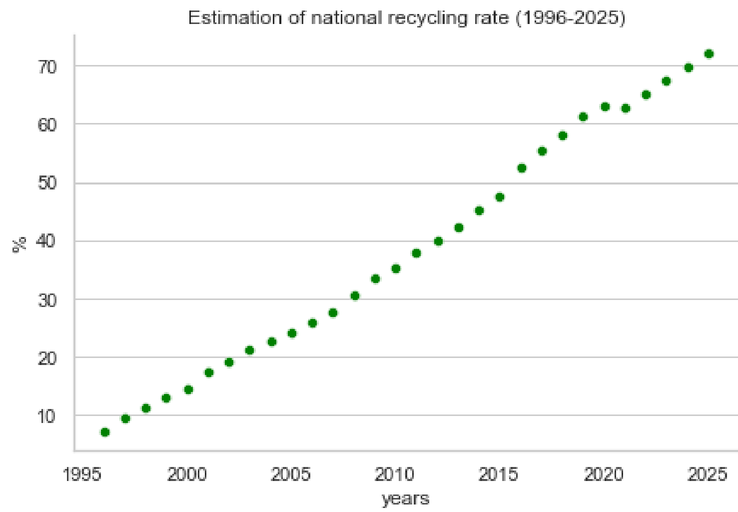

Figure 36: National waste production per capita trend

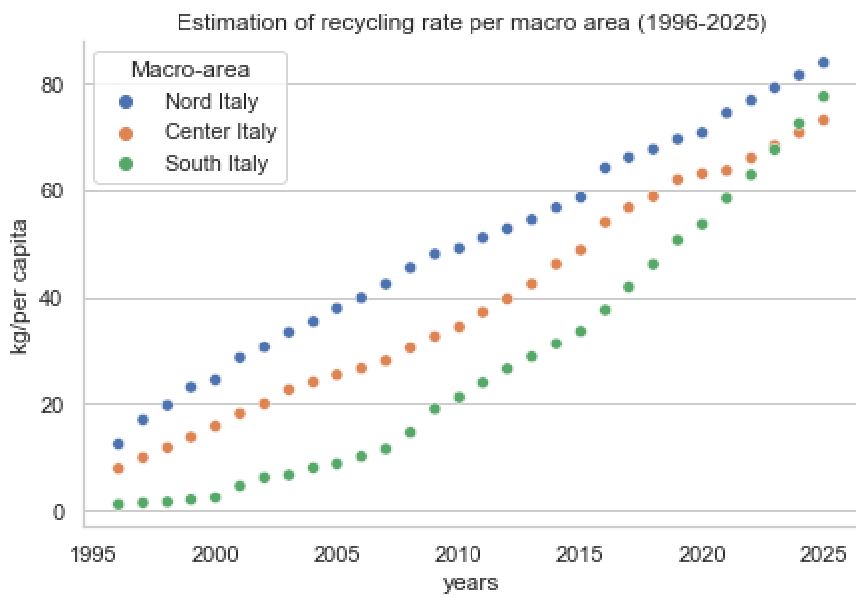

Figure 37: Macro-regional recycling rate future trend

possible to see how the rate has increase at an exponential rate in the last few years (from 2015).

```
Model coefficients are: [ 0.00000000e+00  1.23456731e+03 -3.07325370e-01]
Mean Absolute Error, MAE is  12.524061765277565
Mean Squared Error, MSE is  202.55834980359458
Root Mean Square Error is  14.23229952620428
Coefficient of determination or R squared value is  0.7201189972289177
```

Figure 38: Model coefficients and errors for North Italy

```
Model coefficients are: [ 0.00000000e+00  2.64861609e+03 -6.59573085e-01]
Mean Absolute Error, MAE is  17.7003650918635
Mean Squared Error, MSE is  397.0559153399963
Root Mean Square Error is  19.926261951003163
Coefficient of determination or R squared value is  0.7586789571609389
```

Figure 39: Model coefficients and errors for Center Italy

```
Model coefficients are: [ 0.00000000e+00  1.61526386e+03 -4.02262403e-01]
Mean Absolute Error, MAE is  7.552503971910687
Mean Squared Error, MSE is  133.5068787821111
Root Mean Square Error is  11.55451767847153
Coefficient of determination or R squared value is  0.7735984001450235
```

Figure 40: Model coefficients and errors for South Italy

```
Veneto
Model coefficients are: [ 0.00000000e+00  3.13687365e+02 -7.74478910e-02]
Mean Absolute Error, MAE is  0.9730342965773481
Mean Squared Error, MSE is  1.2166710505172855
Root Mean Square Error is  1.1030281277090288
Coefficient of determination or R squared value is  0.9971304405320705
```

Figure 41: Recycling rate and errors for Region Veneto

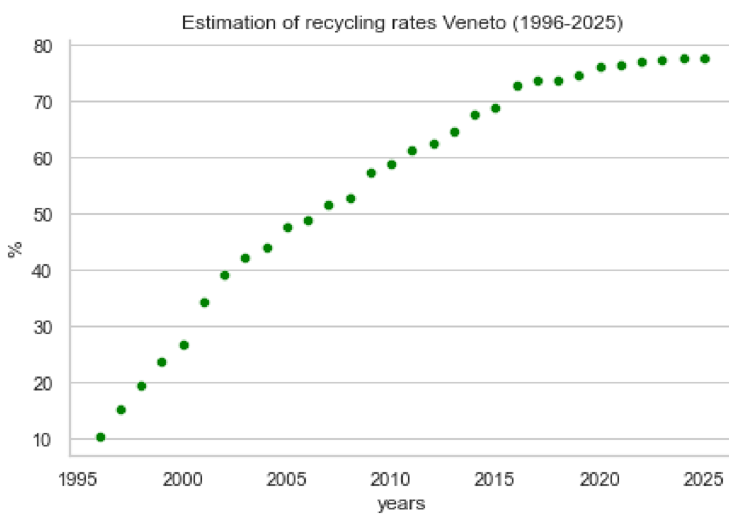

Figure 42: Recycling rate trend for Region Veneto

```
Sicilia
Model coefficients are: [ 0.00000000e+00  9.82104240e+04 -4.89968132e+01  8.14812617e-03]
Mean Absolute Error, MAE is  1.3071079397314618
Mean Squared Error, MSE is  2.729773489544008
Root Mean Square Error is  1.6522026175817563
Coefficient of determination or R squared value is  0.9866433973981287
```

Figure 43: Recycling rate and errors for Region Sicily

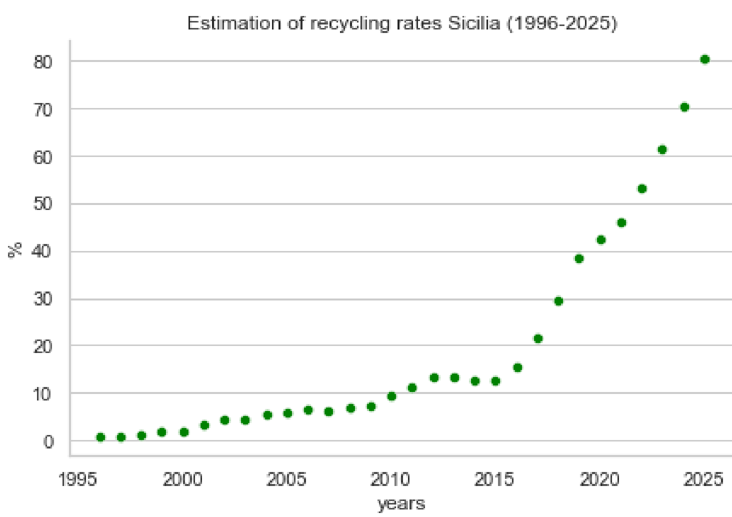

Figure 44: Recycling rate trend for Region Sicily

4.4 Deep Learning: fourth level of analysis

As mentioned in Chapter 3, the task of image classification is performed on a dataset of garbage images, whose composition can be seen in Figure 45. The class that has the highest number of images is the *clothes*, accounting for one third of the dataset.

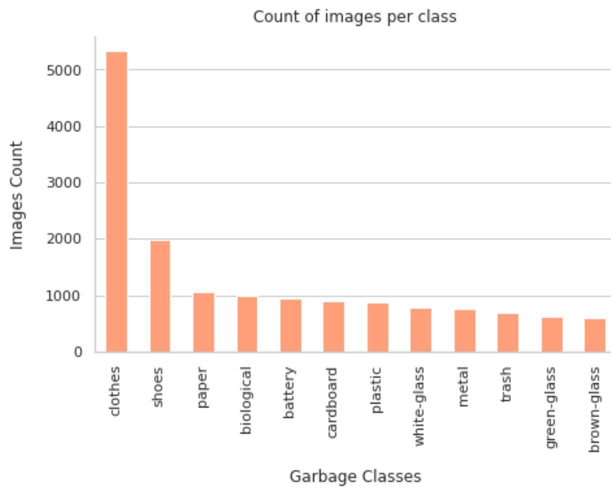

Figure 45: Images per class of the Garbage Collection dataset

The different experiments carried out have been summarized in the following table.

name	value
model	MobileNet, NetV2B0
batch size	32, 64, 128
epochs	early stopping
transfer learning	None, ImageNet

Table 4: Parameters tested on the models

All the algorithms have been evaluated through accuracy on the test set, that resulted in being always above 90%.

In particular the best result achieved by the first model (MobileNet) was achieved trough the Transfer Learning of weights from ImageNet, with a batch size of 128 images, with an accuracy of 94.2% after only 6 epochs. In Figure 46 a summary of the model and its loss and accuracy on the training and validation set.

Note that on top of the model an additional layer has been added to perform classify the dataset in 12 classes. This is the only layer that has been trained during the training phase, since Transfer Learning technique was used (12,300 parameters were trained).

```
Model: "MobileNet"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
preprocessing (Lambda)       (None, 320, 320, 3)       0
_________________________________________________________________
mobilenet_1.00_224 (Function (None, 10, 10, 1024)      3228864
_________________________________________________________________
global_average_pooling2d_2 ( (None, 1024)              0
_________________________________________________________________
softmax_classification (Dens (None, 12)                12300
=================================================================
Total params: 3,241,164
Trainable params: 12,300
Non-trainable params: 3,228,864
_________________________________________________________________
```

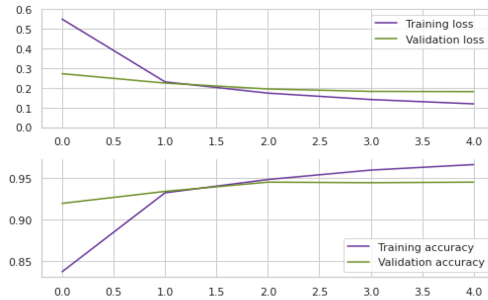

Figure 46: MobileNet model

The second model used is DenseNet121. The best performance was achieved, this time, without transfer learning techniques: the accuracy on the test-set is of 95.81% and the model only took 8 epochs to be trained. In Figure 47 a summary of the model and its loss and accuracy on the training and validation set.

Note that on top of the model an additional layer has been added to perform classify the dataset in 12 classes. This is the only layer that has been trained during the training phase, since Transfer Learning technique was used (12,300 parameters were trained).

```
Model: "DenseNet"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lambda_3 (Lambda)            (None, 320, 320, 3)       0
_________________________________________________________________
densenet121 (Functional)     (None, 10, 10, 1024)      7037504
_________________________________________________________________
global_average_pooling2d_3 ( (None, 1024)              0
_________________________________________________________________
dense_3 (Dense)              (None, 12)                12300
=================================================================
Total params: 7,049,804
Trainable params: 12,300
Non-trainable params: 7,037,504
_________________________________________________________________
```

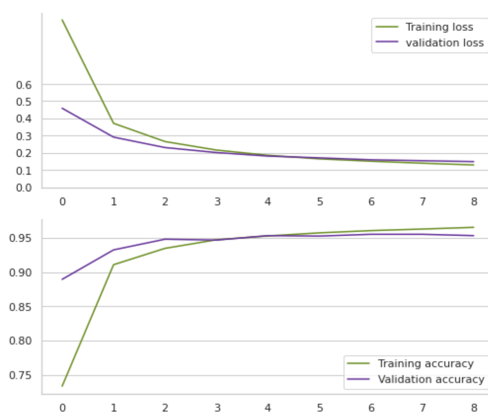

Figure 47: DenseNet121 model

Chapter 5

Recommendations

The analysis conducted on the dataset could be expanded to each regional and provincial scenario in order to identify which areas of Italy needs a targeted intervention.

In those regions where the waste production per capita is higher (North Italy), awareness campaigns could be planned by local governments, such as (LINDANET et al. 2022). Through the prediction and estimation of solid waste it is possible to improve municipal solid waste management plans, even if the models of prediction would have to include more economic and socio-demographic data, to cover as much uncertainties as possible (Kolekar et al. 2016).

Moreover, in those municipalities where the waste sorting rates are low (South Italy and Liguria), Smart Bins could be introduced in houses, offices or in replacement of urban waste bins to help citizen in the segregation process. In particular, this kind of technologies could improve the performances of those municipalities affected by constraints imposed by geographic conformation, socio-economic factors and climate, such as mountain areas, where waste management in mountain areas requires more efforts and is more expensive compared to other areas (European Commission 2000).

Ultimately, the study conducted could support the realization of reports such as the Municipal Waste Compliance Promotion Exercise proposed by the European Union in 2014-2015. The Commission released factsheets and roadmaps for 10 European nations summarizing the conditions of the country and suggesting ways to improve it, such as the documents proposed by the EU for South Italy area, since, as shown in the data, its performances in terms of waste management are not yet aligned with the European requirements (European Commission 2011*a*), (European Commission 2011*b*). The EU then required each Member States to produce a report illustrating the situation of waste management in the country.

Chapter 6

Conclusion and Future Works

6.1 Conclusion

This work considered data regarding the amount of solid waste produced by every municipality for more than 16 different categories of waste in Italy in 2020, collected and published by the Italian Institute for Environmental Protection and Research, ISPRA. The analysis conducted looked in depth at the waste production, recycling rates and waste composition at a macro-regional, regional and provincial level trying to identify patterns and unearth peculiar scenarios, like the case of Liguria's provinces recycling rates described in Section 4.1.1 or the relation between the GDP and the sorted waste per capita proven in Section 4.2.2. After carrying out descriptive and inferential statistics analysis, machine learning was applied to the dataset to predict waste generation per capita and national recycling rate. To conclude, CNN models were used to classify an additional dataset of garbage to perform Automated Waste Segregation, in the context of Smart Bins as smart devices improving the recycling process in households, reaching an accuracy of 94% and 95%.

6.2 Future Work

This report represents a case study on the Italian situation, but the methodology underneath could be applied to any country to understand the performances and identify the lacks of its recycling rates. From an exhaustive study of the waste production and composition is possible to develop and propose tailored solutions to support the waste management process and ultimately, reduce the demand of new resources.

To improve the quality of the waste segregation performed in Section 4.4 a wider and more various dataset would be necessary. The dataset used classifies waste into

12 categories, but the images are mostly about textile waste and not heterogeneous enough. In addition, the algorithm identifies only one object at a time, while a future improvement could be the detection and classification of multiple discarded objects.

Datasets such as TrashNet (a public benchmark in waste classification), WasteNet, ComposNet and individual dataset present on platform such as Kaggle or GitHub could be combined to create a broad dataset or data augmentation techniques such as Generative Adversarial Networks (GAN) could be applied (Alsabei et al. 2021). The algorithms tested could be compared with other models of CNN and eventually implemented in Smart Bins only IoT based such as the work proposed by Ali et al. (2020), Tambekar et al. (2018) and Haque et al. (2020) or improve the performances of other Smart Bins by Sunny et al. (2019), Aazam et al. (2016) and Wijaya et al. (2017) with a less energy consuming algorithm.

Bibliography

Aazam, M., St-Hilaire, M., Lung, C.-H. & Lambadaris, I. (2016), Cloud-based smart waste management for smart cities, *in* '2016 IEEE 21st international workshop on computer aided modelling and design of communication links and networks (CAMAD)', IEEE, pp. 188–193.

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y. & Zheng, X. (2015), 'TensorFlow: Large-scale machine learning on heterogeneous systems'. Software available from tensorflow.org.
URL: *https://www.tensorflow.org/*

Ahmad, K., Khan, K. & Al-Fuqaha, A. (2020), 'Intelligent fusion of deep features for improved waste classification', *IEEE access* **8**, 96495–96504.

Ali, T., Irfan, M., Alwadie, A. S. & Glowacz, A. (2020), 'Iot-based smart waste bin monitoring and municipal solid waste management system for smart cities', *Arabian Journal for Science and Engineering* **45**(12), 10185–10198.

Alsabei, A., Alsayed, A., Alzahrani, M. & Al-Shareef, S. (2021), 'Waste classification by fine-tuning pre-trained cnn and gan', *International Journal of Computer Science & Network Security* **21**(8), 65–70.

Bandara, N. J., Hettiaratchi, J. P. A., Wirasinghe, S. & Pilapiiya, S. (2007), 'Relation of waste generation and composition to socio-economic factors: a case study', *Environmental monitoring and assessment* **135**(1), 31–39.

Bobulski, J. & Kubanek, M. (2019), Cnn use for plastic garbage classification method, *in* 'Proceedings of the 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining'.

Chen, D. M.-C., Bodirsky, B. L., Krueger, T., Mishra, A. & Popp, A. (2020), 'The world's growing municipal solid waste: Trends and impacts', *Environmental Research Letters* **15**(7), 074021.

Chen, J., Lu, W. & Xue, F. (2021), '"looking beneath the surface": A visual-physical feature hybrid approach for unattended gauging of construction waste composition', *Journal of Environmental Management* **286**, 112233.

Chollet, F. et al. (2015), 'Keras'.
URL: *https://github.com/fchollet/keras*

Chu, Y., Huang, C., Xie, X., Tan, B., Kamal, S. & Xiong, X. (2018), 'Multilayer hybrid deep-learning method for waste classification and recycling', *Computational Intelligence and Neuroscience* **2018**.

Dieguez, T. (2020), Operationalization of circular economy: A conceptual model, *in* 'Handbook of Research on Entrepreneurship Development and Opportunities in Circular Economy', IGI Global, pp. 38–60.

Doble, M. & Kumar, A. (2005), Chapter 15 - waste from nuclear plants, *in* M. Doble & A. Kumar, eds, 'Biotreatment of Industrial Effluents', Butterworth-Heinemann, Burlington, pp. 169–175.
URL: *https://www.sciencedirect.com/science/article/pii/B9780750678384500166*

EU (2008), 'Directive 2008/98/ec of the european parliament and of the council of 19 november 2008 on waste and repealing certain directives'. (OJ L 312, 22.11.2008, p. 3-30).

EU (2018a), 'Directive (eu) 2018/851 of the european parliament and of the council of 30 may 2018 amending directive 2008/98/ec on waste'. (OJ L 150, 14.06.2018, p. 109-140).

European Commission (2000), 'Guida della gestione dei rifiuti in aree di montagna'.

European Commission (2008), 'Waste framework directive'.
URL: *https://ec.europa.eu/environment/topics/waste-and-recycling/waste-framework-directive*$_en$

European Commission (2011*a*), 'South italy – factsheet'.
URL: *https://ec.europa.eu/environment/pdf/waste/framework/IT_SOUTH_factsheet_FINAL.pdf*

European Commission (2011*b*), 'South italy – roadmap'.
URL: *https://ec.europa.eu/environment/pdf/waste/framework/IT_SOUTH_Roadmap_FINAL.pdf*

European Commission (2020), *Circular economy action plan : for a cleaner and more competitive Europe*, Publications Office.

European Standards (2012), 'En 16403:2012 (waste management - waste visual elements)'. (20G00135).
URL: *https://www.en-standard.eu/12-30258346-dc-bs-en-16403-waste*

Foundation, E. M. (2015), 'Circular economy'.
URL: *https://ellenmacarthurfoundation.org/topics/circular-economy-introduction/overview/*

Gazzetta Ufficiale (2006), 'Norme in materia ambientale'. (n°152).

Gazzetta Ufficiale (2008), 'Disposizioni urgenti per lo sviluppo economico, la semplificazione, la competitività, la stabilizzazione della finanza pubblica e la perequazione tributaria'. (n°112).

Gazzetta Ufficiale (2020), 'Attuazione della direttiva (ue) 2018/851 che modifica la direttiva 2008/98/ce relativa ai rifiuti e attuazione della direttiva (ue) 2018/852 che modifica la direttiva 1994/62/ce sugli imballaggi e i rifiuti di imballaggio'. (20G00135).

Gothai, E., Thamilselvan, R., Natesan, P., Keerthivasan, M., Kabinesh, K. & Ruban, D. K. (2022), Plastic waste classification using cnn for supporting 3r's principle, *in* '2022 International Conference on Computer Communication and Informatics (ICCCI)', IEEE, pp. 01–07.

Gutberlet, J. & Uddin, S. M. N. (2017), 'Household waste and health risks affecting waste pickers and the environment in low-and middle-income countries', *International journal of occupational and environmental health* **23**(4), 299–310.

Gyawali, D., Regmi, A., Shakya, A., Gautam, A. & Shrestha, S. (2020), 'Comparative analysis of multiple deep cnn models for waste classification', *arXiv preprint arXiv:2004.02168* .

Haque, K. F., Zabin, R., Yelamarthi, K., Yanambaka, P. & Abdelgawad, A. (2020), An iot based efficient waste collection system with smart bins, *in* '2020 IEEE 6th World Forum on Internet of Things (WF-IoT)', IEEE, pp. 1–5.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. & Adam, H. (2017), 'Mobilenets: Efficient convolutional neural networks for mobile vision applications', *arXiv preprint arXiv:1704.04861* .

Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. (2017), Densely connected convolutional networks, *in* 'Proceedings of the IEEE conference on computer vision and pattern recognition', pp. 4700–4708.

ISPRA (2021), 'Catasto nazionale rifiuti'.
URL: *https://www.catasto-rifiuti.isprambiente.it/index.php?pg=downloadTari*

Iyamu, H., Anda, M. & Ho, G. (2020), 'A review of municipal solid waste management in the bric and high-income countries: A thematic framework for low-income countries', *Habitat International* **95**, 102097.
URL: *https://www.sciencedirect.com/science/article/pii/S0197397519303467*

Kaza, S., Yao, L., Bhada-Tata, P. & Van Woerden, F. (2018), *What a waste 2.0: a global snapshot of solid waste management to 2050*, World Bank Publications.

Kolekar, K., Hazra, T. & Chakrabarty, S. (2016), 'A review on prediction of municipal solid waste generation models', *Procedia Environmental Sciences* **35**, 238–244.

Legambiente & Liguria, R. (2021), 'Comuni ricicloni liguria 2020'.
URL: *http://www.ricicloni.it/assets/edizioni-regionali/liguria-2020.pdf*

Liang, S. & Gu, Y. (2021), 'A deep convolutional neural network to simultaneously localize and recognize waste types in images', *Waste Management* **126**, 247–257.

Lin, K., Zhao, Y., Kuo, J.-H., Deng, H., Cui, F., Zhang, Z., Zhang, M., Zhao, C., Gao, X., Zhou, T. et al. (2022), 'Toward smarter management and recovery of municipal solid waste: A critical review on deep learning approaches', *Journal of Cleaner Production* p. 130943.

LINDANET, PLASTECO, BIOREGIO, CircPro & CONDEREFF (2022), 'Waste management communication task force'.
URL: *https://www.interregeurope.eu/good-practices/waste-management-communication-task-force*

Magazzino, C., Mele, M., Schneider, N. & Sarkodie, S. A. (2021), 'Waste generation, wealth and ghg emissions from the waste sector: Is denmark on the path towards circular economy?', *Science of The Total Environment* **755**, 142510.
URL: *https://www.sciencedirect.com/science/article/pii/S0048969720360393*

Mao, W.-L., Chen, W.-C., Wang, C.-T. & Lin, Y.-H. (2021), 'Recycling waste classification using optimized convolutional neural network', *Resources, Conservation and Recycling* **164**, 105132.

Mazzanti, M. & Zoboli, R. (2008), 'Waste generation, waste disposal and policy effectiveness: Evidence on decoupling from the european union', *Resources, Conservation and Recycling* **52**(10), 1221–1234.
URL: *https://www.sciencedirect.com/science/article/pii/S0921344908001079*

Milonton, J. (2021), 'Linear vs. circular, history and context'.
URL: *https://edmire.design/blog-x/circular-economy-history-and-context*

Minelgaitė, A. & Liobikienė, G. (2019), 'The problem of not waste sorting behaviour, comparison of waste sorters and non-sorters in european union: Cross-cultural analysis', *Science of the Total Environment* **672**, 174–182.

Mukherjee, A. G., Wanjari, U. R., Chakraborty, R., Renu, K., Vellingiri, B., George, A., CR, S. R. & Gopalakrishnan, A. V. (2021), 'A review on modern and smart technologies for efficient waste disposal and management', *Journal of Environmental Management* **297**, 113347.

Nanda, S. & Berruti, F. (2021), 'Municipal solid waste management and landfilling technologies: a review', *Environmental Chemistry Letters* **19**(2), 1433–1456.

Nestlé (2021), 'In italia più di 200 modalità di raccolta differenziata'.
URL: *https://www.nestle.it/media/pressreleases/allpressreleases/italia-200-modalità-raccolta-differenziata*

Ozcan, H. K., Guvenc, S. Y., Guvenc, L. & Demir, G. (2016), 'Municipal solid waste characterization according to different income levels: A case study', *Sustainability* **8**(10).
URL: *https://www.mdpi.com/2071-1050/8/10/1044*

Periathamby, A. (2011), Chapter 8 - municipal waste management, *in* T. M. Letcher & D. A. Vallero, eds, 'Waste', Academic Press, Boston, pp. 109–125.
URL: *https://www.sciencedirect.com/science/article/pii/B9780123814753100087*

Rahman, M. W., Islam, R., Hasan, A., Bithi, N. I., Hasan, M. M. & Rahman, M. M. (2020), 'Intelligent waste management system using deep learning with iot', *Journal of King Saud University-Computer and Information Sciences* .

Ritchie, H., Roser, M. & Rosado, P. (2020), 'Emissions by sector'.
URL: *https://ourworldindata.org/emissions-by-sector*

Ruiz, V., Sánchez, Á., Vélez, J. F. & Raducanu, B. (2019), Automatic image-based waste classification, *in* 'International Work-Conference on the Interplay Between Natural and Artificial Computation', Springer, pp. 422–431.

Sallang, N. C. A., Islam, M. T., Islam, M. S. & Arshad, H. (2021), 'A cnn-based smart waste management system using tensorflow lite and lora-gps shield in internet of things environment', *IEEE Access* **9**, 153560–153574.

Sivakumar, K. & Sugirtharan, M. (2010), 'Impact of family income and size on per capita solid waste generation: a case study in manmunai north divisional secretariat division of batticaloa'.

Soundarya, B., Parkavi, K., Sharmila, A., Kokiladevi, R., Dharani, M. & Krishnaraj, R. (2022), Cnn based smart bin for waste management, *in* '2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT)', IEEE, pp. 1405–1409.

Srinilta, C. & Kanharattanachai, S. (2019), Municipal solid waste segregation with cnn, *in* '2019 5th International Conference on Engineering, Applied Sciences and Technology (ICEAST)', IEEE, pp. 1–4.

Sunny, M. S. H., Dipta, D. R., Hossain, S., Faruque, H. M. R. & Hossain, E. (2019), Design of a convolutional neural network based smart waste disposal system, *in* '2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)', IEEE, pp. 1–5.

Symbola (2021).
URL: *https://www.symbola.net/ricerca/leconomia-circolare-in-italia-per-il-next-generation-eu/*

Tambekar, A., Channe, V., Raut, A., Chahodkar, A., Bhoskar, A. & Thool, A. (2018), 'Innovation waste collection system using wireless sensor network aka 'smart dustbin'', *Innovation* **5**(02).

United Nations (2015), 'Transforming our world: The 2030 agenda for sustainable development'.
URL: *https://sdgs.un.org/2030agenda*

Wang, Y., Zhao, W. J., Xu, J. & Hong, R. (2020), 'Recyclable waste identification using cnn image recognition and gaussian clustering', *arXiv preprint arXiv:2011.01353* .

White, G., Cabrera, C., Palade, A., Li, F. & Clarke, S. (2020), 'Wastenet: Waste classification at the edge for smart bins', *arXiv preprint arXiv:2006.05873* .

Wijaya, A. S., Zainuddin, Z. & Niswar, M. (2017), Design a smart waste bin for smart waste management, *in* '2017 5th International Conference on Instrumentation, Control, and Automation (ICA)', IEEE, pp. 62–66.

Wilson, D. C. & Velis, C. A. (2015), 'Waste management–still a global challenge in the 21st century: An evidence-based call for action'.

Appendix A

Appendix

A.1 Methodology

```python
# Loading the full dataset to inquiry at regional level
dataset_2020 = pd.read_excel('production/1646058903_Anno_2020_RUComunali.xlsx',
    na_values = '-')
# Cleaning the dataset first column
dataset_2020['IstatComune'].replace('\t')
dataset_2020.head()
dataset_2020.describe()
# Number of NaN
print(dataset_2020.isna().sum()/dataset_2020.shape[0]*100)
print(dataset_2020.isna().sum())
```

Listing A.1: Loading the dataset

```python
#Replacing the missing values with the values of the same region, where the number of
    inhabitants is the most similar (the less difference)
attributes = dataset_2020.columns[6:-5]
attributes = attributes.append(dataset_2020.columns[20:24])
attributes = attributes.drop( "Totale RD (t)")
attributes = attributes.drop( "Totale RU (t)")
print(list(attributes))

for a in attributes:
    print()
    print(a.upper())
    #dataset_2020.to_excel('production/Anno_2020_RUComunali_Complete.xlsx')
    for index, row in dataset_2020.iterrows():
        if row["Dato riferito a"].startswith("Vedi agg"):
            continue
        value = row[a]
```

```
        #If the value is not null, then skip; otherwise sobstitute it
        #print("Elaborating for", row["Comune"], "whose value for Totale RU (t) is", value, end = ". ")
        if math.isnan(value):
            #print("Comune", row["Comune"])
            #print("Popolazione", row["Popolazione"])
            #print("Attribute: ", a, row[a])
            village_population_delta = 0
            village_value = 0
            # Look for the list of villages that in the same region have the same population and are not null
            #Flag to check the first iteration
            flag = True
            for i, r in dataset_2020.iterrows():
                # Check if the same region
                if row["Regione"] == r["Regione"]:
                    if not math.isnan(r[a]):
                        #clear_output(wait=True)
                        population = row["Popolazione"]
                        #print("Other village:", end = " ")

                        #Keep the one with less difference in population
                        #print(r["Comune"], "Delta di popolazione:", abs(population - r["Popolazione"]))

                        #If the difference is lower, keep the one with less difference
                        if flag == True or abs(row["Popolazione"] - r["Popolazione"]) < village_population_delta:
                            #print(r["Comune"], "Delta di popolazione:", abs(population - r["Popolazione"]))
                            #print("Selected!")
                            village_population_delta = abs(row["Popolazione"] - r["Popolazione"])
                            village_value = r[a]
                            village_pop = r["Popolazione"]
                            village_value_norm = r[a] * row["Popolazione"] / r["Popolazione"]
                            flag = False
            dataset_2020.loc[index, a] = village_value_norm
            #print("*******************************")
            #print("Sobstituted value", village_value_norm)
            #print(village_value, ":", village_pop, "= X :",  row["Popolazione"])
            #print("*******************************")
            print(".", end = "")
dataset_2020.isna().sum()
```

Listing A.2: Replacing missing values

```python
attributes_a = []
attributes_a.append(dataset_2020.columns[7:-5])
attributes_a = list(attributes_a[0])
print(attributes_a)
add = list(dataset_2020.columns[-4:-2])
attributes_b = attributes_a + add
print(attributes_b)
for index, row in dataset_2020.iterrows():
    if row["Dato riferito a"].startswith("Vedi agg"):
        continue
    ru = row["Totale RU (t)"]
    rd = row["Totale RD (t)"]
    rr = row["Percentuale RD (%)"]
    if math.isnan(ru):
        value = 0
        for a in list(attributes_b):
            value = value + row[a]
        #print(value)
        dataset_2020.loc[index, "Totale RU (t)"] = value
    if math.isnan(rd):
        value = 0
        for a in list(attributes_a):
            value = value + row[a]
        #print(value)
        dataset_2020.loc[index, "Totale RD (t)"] = value

for index, row in dataset_2020.iterrows():
    if row["Dato riferito a"].startswith("Vedi agg"):
        continue
    ru = row["Totale RU (t)"]
    rd = row["Totale RD (t)"]
    rr = row["Percentuale RD (%)"]
    if math.isnan(rr) and not math.isnan(rd) and not math.isnan(ru):
        if rd/ru > 1.0:
            dataset_2020.loc[index, "Percentuale RD (%)"] = 1.0
        else:
            dataset_2020.loc[index, "Percentuale RD (%)"] = rd/ru
dataset_2020.isna().sum()
```

Listing A.3: Replacing missing values

A.2 Results

A.3 Descriptive statistics

A.3.1 Recycling rates

```
import pandas as pd
import numpy as np
import matplotlib
import matplotlib.pyplot as plt
import seaborn as sns
import math
dataset_macro_2020 = pd.read_excel('production/Anno_2020_RDFrazioniItalia.xlsx')
sns.set(rc={'figure.figsize':(7.2,4.45)})
sns.set(style="whitegrid")
rd = dataset_macro_2020[dataset_macro_2020["Frazione Merceologica"] == "TOTALE RD"]["
    Italia (t)"]
total = dataset_macro_2020[dataset_macro_2020["Frazione Merceologica"] == "TOTALE RU"
    ]["Italia (t)"]
rd = float(rd)/float(total)
ri = 1-rd
labels = 'Recycling', 'Unsorted waste'
sizes = [rd, ri]
fig1, ax1 = plt.subplots()
ax1.pie(sizes, labels=labels, colors = ("olivedrab", "rebeccapurple"), autopct='%1.1f
    %%', startangle=90)
ax1.axis('equal')  # Equal aspect ratio ensures that pie is drawn as a circle.
ax1.set_title('Recycling rate of Italy in 2020', fontsize = 13)
plt.show()
```

Listing A.4: Italian recycling rate

```
sns.set(rc={'figure.figsize':(7.2,4.45), 'font.size': 13})
sns.set(style="whitegrid")

rd = dataset_macro_2020[dataset_macro_2020["Frazione Merceologica"] == "TOTALE RD"]
total = dataset_macro_2020[dataset_macro_2020["Frazione Merceologica"] == "TOTALE RU"]
plot = rd.iloc[:, 1:].values/total.iloc[:, 1:].values*100
labels = ["North", "Center", "South", "Italy"]
values = []
for p in plot[0]: values.append("{0:.2f}".format(p) + str("%"))

fig, ax = plt.subplots()
bar = ax.bar(labels, plot[0], color = "olivedrab")
ax.set_ylim([0, 100])
```

```
ax.bar_label(bar, values, fmt='0:.2f', padding=1, fontsize = 13)
ax.set_title('Recycling rate of Italy in 2020', fontsize = 13)
sns.despine()
plt.grid(axis = "x")
plt.axhline(y=50, color='darkblue', linestyle='--', linewidth=2, zorder=-1)
t = plt.text(1.8, 45, 'European average', fontsize=13, va='center', backgroundcolor="w
    ", color='darkblue')
t.set_bbox(dict(facecolor='w', alpha=0, edgecolor='w'))
plt.show()
```

Listing A.5: Italian recycling rate

```
dataset_2020 = pd.read_excel("production/Anno_2020_RUComunali_Complete.xlsx",
    na_values = "-")
# Plot showing each regional value of percentage of recycling (from summation of total
     waste)

df = dataset_2020.groupby("Regione").sum()["Totale RD (t)"]/dataset_2020.groupby("
    Regione").sum()["Totale RU (t)"]*100

sns.set(rc={'figure.figsize':(14.4,4.45)})
sns.set(rc={'font.size': 13})
sns.set(style="whitegrid")

fig, ax = plt.subplots()
df.plot.bar(color = color_12, ax = ax)
ax.set_title('Regional waste sorting rate in 2020')
ax.set_xlabel("")
ax.set_ylim([0, 100])
sns.despine()

for bar in ax.patches:
    ax.annotate(str(format(bar.get_height(), '.2f')+"%"),
                   (bar.get_x() + bar.get_width() / 2,
                    bar.get_height()), ha='center', va='center',
                   size=10, xytext=(0, 10),
                   textcoords='offset points')

plt.grid(axis = "x")
plt.axhline(y=np.mean(df), color='olivedrab', linestyle='--', linewidth=2, zorder=-1,
    alpha=0.4)
plt.show()
```

Listing A.6: Regional recycling rates

```
df1 = dataset_2020.groupby("Regione").sum()["Totale RD (t)"]/dataset_2020.groupby("
    Regione").sum()["Totale RU (t)"]*100
```

```
print(df1.describe())
df2 = dataset_2020.groupby("Regione").describe()["Percentuale RD (%)"]["mean"]*100
df2.describe()
df3 = pd.DataFrame({"Ratio of sorted waste": df1, "Average of sorted waste": df2})
sns.set(rc={'font.size': 13})
sns.set(rc={'figure.figsize':(7.2,4.45)})
sns.set(style="whitegrid")
fig, ax = plt.subplots()
sns.boxplot(data = df3, palette = "Paired", ax = ax)
ax.set_title('Regional recycling rate distribution in 2020')
ax.set_ylabel("%")
ax.set_ylim([40, 80])
sns.despine()
plt.show()
```

Listing A.7: Boxplot regional recycling rates

```
sns.set(rc={'figure.figsize':(7.2,4.45)})
sns.set(style="whitegrid")
sns.despine()
df = dataset_2020[dataset_2020["Regione"] == "Liguria"]
mean = np.mean(df)["Percentuale RD (%)"]
# Considering the region Liguria barchar
ax = sns.displot(data=df, x="Percentuale RD (%)", col="Regione", stat="percent", kde=
    True, color = "olivedrab", height=4.45, aspect=7.2/4.45)
ax.set_titles('Distribution of region Liguria recycling rate in 2020', size = 13)
plt.axvline(mean, linewidth=2, linestyle="--", color='olivedrab', ymax=0.9, alpha =
    0.7)
plt.grid(axis = "x")
plt.xlim(0, 1)
plt.show()
```

Listing A.8: Region Liguria recycling rate

```
from scipy.stats import norm
import scipy.stats as stats
import pylab

sns.set(rc={'figure.figsize':(7.2,4.45)})
sns.set(style="whitegrid")
df = dataset_2020[dataset_2020["Regione"] == "Liguria"]
fig, ax = plt.subplots()
stats.probplot(df["Percentuale RD (%)"], dist="norm", plot=ax)
sns.despine()
ax.set_title('Quantile-Quantile plot of region Liguria recycling rate in 2020')
ax.set_xlabel("")
ax.set_ylabel("")
```

```
ax.get_lines()[0].set_color('olivedrab')
ax.get_lines()[0].set_marker('.')
ax.get_lines()[0].set_markersize(4)
ax.get_lines()[1].set_linewidth(2)
ax.get_lines()[1].set_color('darkred')
plt.show()
```

Listing A.9: QQPlot of Region Liguria recycling rates

```
import matplotlib as mpl
sns.set(rc={'font.size': 13})
sns.set(rc={'figure.figsize':(7.2,4.45)})
sns.set(style="whitegrid")
data = df.groupby("Provincia").describe()["Percentuale RD (%)"]["mean"]*100
fig, ax = plt.subplots()
plt.tight_layout()
ax = data.plot.barh(color = "olivedrab", alpha = 0.7)
ax.set_title('Provinces of region Liguria recycling rates in 2020')
ax.set_ylabel("")
ax.set_xlabel("Percentuale RD (%)")
sns.despine()
plt.grid(axis = "y")
plt.axvline(x=np.mean(data), color="olivedrab", linestyle='--', linewidth=2, zorder
    =1,)
totals = []
# find the values and append to list
for i in ax.patches:
    totals.append(i.get_width())
# set individual bar lables using above list
total = sum(totals)
# set individual bar lables using above list
for i in ax.patches:
    # get_width pulls left or right; get_y pushes up or down
    ax.text(i.get_width()+.3, i.get_y()+i.get_height()/2.5, str(round(i.get_width(),
    2)) +'%', fontsize=12)
plt.show()
```

Listing A.10: Region Liguria' provinces recycling rates

```
from shapely.geometry import Point
import requests
sns.set(rc={'font.size': 13})
def percentageToHsl(percentage, hue0, hue1):
    hue = (percentage * (hue1 - hue0)) + hue0;
    return 'hsl(' + str(hue) + ', 100%, 50%)';
points = []
values = []
```

```
names = []
for index, row in df.iterrows():
    url = "https://nominatim.openstreetmap.org/search?q=" + row["Comune"] + "+Liguria&
    format=geojson"
    response = requests.get(url)
    lat = response.json()["features"][0]["geometry"]["coordinates"][1]
    lon = response.json()["features"][0]["geometry"]["coordinates"][0]
    point = Point(lon, lat)
    points.append(point)
    names.append(row["Comune"])
    values.append(row["Percentuale RD (%)"])

d = {'value': values, 'geometry': points, 'name': names}
```

Listing A.11: Pre-processing data for plotting a map of Region Liguria's municipalities recycling rates

```
import geopandas as gpd

liguria = gpd.read_file("maps/regions-with-provinces/liguria/liguria.shp")
sns.set(rc={'figure.figsize':(10,5)})
sns.set(style="whitegrid")
fig, ax = plt.subplots()
liguria.plot(color = "tan", ax = ax)
ax.set_title("Municipalities of region Liguria recycling rate in 2020")
ax.get_xaxis().set_ticks([])
ax.get_yaxis().set_ticks([])
plt.grid(axis = "both")
data = gpd.GeoDataFrame(d)
data = data.set_crs(crs = "epsg:4326")
data = data.to_crs(liguria.crs)
data.plot(ax = ax, column = "value", cmap = "summer_r", alpha = 0.6)
for idx, row in data.iterrows():
    if row['name'] == "SAVONA" or row['name'] == "LA SPEZIA" or row['name'] == "
    IMPERIA" or row['name'] == "GENOVA":
        plt.annotate(text=row['name'].title(), xy=row['geometry'].coords[0],
    horizontalalignment='center')
```

Listing A.12: Region Liguria's municipalities recycling rates

A.3.2 Waste production

```
data = pd.read_excel('production/Anno_2020_ProduzioneRURegionale_Italia.xlsx')
data = data.sort_values("Pro capite RU (kg/ab.*anno)")

```

```
sns.set(rc={'figure.figsize':(20, 8)})
sns.set(style="whitegrid")
fig, ax = plt.subplots()
data.index = data["Provincia"]
ax = data["Pro capite RU (kg/ab.*anno)"].plot.bar(color = "rebeccapurple")
data["Pro capite RD (kg/ab.*anno)"].plot.bar(ax = ax, color = "olivedrab")
ax.set_title('Italian regions\' per capite waste production in 2020', size = 14)
ax.set_xlabel("")
sns.despine()
plt.grid(axis = "x")

plt.axhline(y=488, color='sandybrown', linestyle=(0, (5, 10)), linewidth=2, zorder=-1)
plt.text(0.6, 458, 'Italian average', fontsize=14, va='center', backgroundcolor='w',
    color='sandybrown')
plt.axhline(y=505, color='darkblue', linestyle=(0, (5, 10)), linewidth=2, zorder=-1)
plt.text(0, 535, 'European average', fontsize=14, va='center', backgroundcolor='w',
    color='darkblue')

sns.despine()
plt.show()
```

Listing A.13: Regional solid waste production

```
data = pd.read_excel('production/Anno_2020_ProduzioneRUNazionale.xlsx')
sns.set(rc={'figure.figsize':(7.2, 4.45)})
sns.set(style="whitegrid")
fig, ax = plt.subplots()
data.index = data["Area"]
ax = data["Pro capite RU (kg/ab.*anno)"].plot.bar(color = "rebeccapurple")
data["Pro capite RD (kg/ab.*anno)"].plot.bar(ax = ax, color = "olivedrab")
ax.set_title('Italian regions\' per capite waste production in 2020', size = 13)
ax.set_xlabel("")
sns.despine()
plt.grid(axis = "x")

labels = ["North", "Center", "South", "Italy"]
ax.set_xticklabels(labels)

plt.axhline(y=505, color='darkblue', linestyle=(0, (5, 10)), linewidth=2, zorder=-1)
plt.text(1.5, 475, 'European average', fontsize=14, va='center', backgroundcolor='w',
    color='darkblue')

sns.despine()
plt.show()
```

Listing A.14: National solid waste production

```python
sns.set(rc={'figure.figsize':(20, 8)})
sns.set(rc={'font.size': 13})
sns.set(style="whitegrid")
fig, ax = plt.subplots()
data.index = data["Provincia"]
dt = data["Pro capite RU (kg/ab.*anno)"] - data["Pro capite RD (kg/ab.*anno)"]
dt = dt.sort_values()
ax = dt.plot.bar(color = "rebeccapurple")
ax.set_title('Italian regions\' per capite unsorted waste production in 2020', size =
    14)
ax.set_xlabel("")
sns.despine()

plt.grid(axis = "x")
plt.axhline(y=np.mean(dt), color='olivedrab', linestyle=(0, (5, 10)), linewidth=2,
    zorder=-1)
plt.text(0.6, np.mean(dt)-10, 'Italian average', fontsize=13, va='center',
    backgroundcolor='w', color='olivedrab')
plt.axhline(y=151, color='darkblue', linestyle=(0, (5, 10)), linewidth=2, zorder=-1)
plt.text(0, 140, 'European average', fontsize=14, va='center', backgroundcolor='w',
    color='darkblue')
sns.despine()
plt.show()
```

Listing A.15: National unsorted solid waste production

A.3.3 Waste composition

```python
import squarify

sizes = dataset_macro_2020["Italia (t)"][:-4]
label = dataset_macro_2020["Frazione Merceologica"][:-4]

cmap = matplotlib.cm.Greens
norm = matplotlib.colors.Normalize(vmin=np.min(sizes),vmax=np.max(sizes))

plt.figure(figsize=(16, 6))
title="Tree map of the recycable solid waste composition, size = waste in (t)"
plt.title(title, size=13)

ax = squarify.plot(sizes, label=label, color=color_12, alpha=0.5, pad=True)

plt.axis("off")
plt.show()
```

Listing A.16: Composition of sortable waste

```
sns.set(rc={'figure.figsize':(7.2, 6)})
sns.set(style="whitegrid")

list_ix = list(range(0,13))
df = dataset_macro_2020.iloc[list_ix]

series = dataset_macro_2020.iloc[16, :-1]

df_kg_pc = (df.iloc[:, 1:-1]).div(list(series)[1:], axis = 1)*100

sns.set(rc={'figure.figsize':(7.2, 6)})
sns.set(style="whitegrid")

ax = df_kg_pc.plot.barh(color = ("rebeccapurple", "thistle", "purple"))
ax.set_title('Solid waste composition percentages for macro-area in Italy in 2020')
ax.set_xlabel("%")
ax.set_ylabel("Waste components by type")
plt.grid(axis = "y")
ax.set_yticklabels(df["Frazione Merceologica"])
sns.despine()
plt.savefig('macro_overview_waste_categories_2020.png')
```

Listing A.17: Solid waste composition percentages per macro-area

A.4 Inferential Statistic

```
import matplotlib as mpl
import seaborn as sns

sns.set(rc={'font.size': 13})
sns.set(rc={'figure.figsize':(7.2,6)})
sns.set(style="whitegrid")

fig, ax = plt.subplots()
plt.tight_layout()

ax = data["Pro capite PIL (mil    )"].plot.barh(color = "olivedrab", alpha = 0.7)
ax.set_title('GDP per capita (mil    )')
ax.set_ylabel("")

sns.despine()

plt.grid(axis = "y")

```

```
# set individual bar lables using above list
for i in ax.patches:
    # get_width pulls left or right; get_y pushes up or down
    ax.text(i.get_width()+.3, i.get_y()+i.get_height()/2.5, str(round(i.get_width(),
    2)) +'%', fontsize=12)

plt.show()
```

Listing A.18: Regional GDP per capita

```
data = pd.read_excel('production/Anno_2020_ProduzioneRURegionale_Italia.xlsx')
data = data.drop(columns = ["Istat"], axis = 1)
# https://www.istat.it/it/archivio/237813
data["Pro capite PIL (mil    )"] = [29.4, 36.3, 36.7, 40.7, 31.3, 30.6, 30.4, 33.6,
    30.2, 24.6, 26.2, 32.4, 23.8, 20.2, 18.1, 17.9, 20.9, 16.4, 17.1, 20.1]
correlation = data.corr()
data.index = data["Provincia"]
data = data.sort_values("Pro capite PIL (mil    )")
mask = np.triu(np.ones_like(correlation, dtype=bool))
f, ax = plt.subplots(figsize=(7.2, 4.45))
sns.heatmap(correlation, mask = mask, cmap="PRGn_r", center=0, square=True, linewidths
    =.5, cbar_kws={"shrink": .5}, annot = True, annot_kws = {"size": 11})
```

Listing A.19: Correlation matrix

```
dataset_2020 = pd.read_excel("production/Anno_2020_RUComunali_Complete_Test_2.xlsx",
    na_values = "-")

categories = ["Legno (t)"] #Then performed for the other categories
for c in categories:
    regions = set(dataset_2020["Regione"])
    dataset = {}
    means = []
    for r in regions:
        #print(r)
        data = dataset_2020[dataset_2020["Regione"] == r]
        #print(data["Frazione umida(1) (t)"])
        #print(data["Totale RD (t)"])
        df = data[c].div(list(data["Totale RU (t)"]))*100
        dataset[r] = df[df.isna()==False]
        #means.append(np.mean(dataset[r]))

    dframe = pd.DataFrame.from_dict(dataset)
    #print(dframe)
    data = dframe.describe().loc["mean", :]
    #print(data)

```

```
    fig, ax = plt.subplots()

    ax = data.plot.hist(color = "olivedrab")
    ax.set_title(c)
    ax.set_xlabel("")

    sns.despine()
    plt.grid(axis = "y")
    #print("std: ", np.std(data), "skew: ", data.skew(), "kurtosis:", data.kurt())
    plt.show()
    #fig.savefig('img/regional_prod_organic_2020.png')
```

Listing A.20: T-test on waste categories

```
df_centro = [np.mean(dataset["Toscana"]), np.mean(dataset["Umbria"]), np.mean(dataset["Marche"]), np.mean(dataset["Lazio"])]
df_sud = [np.mean(dataset["Sicilia"]), np.mean(dataset["Abruzzo"]), np.mean(dataset["Campania"]), np.mean(dataset["Sardegna"]),
          np.mean(dataset["Molise"]),
          np.mean(dataset["Basilicata"]),
          np.mean(dataset["Calabria"]),
          np.mean(dataset["Puglia"])]
df_nord = [np.mean(dataset["Veneto"]),
           np.mean(dataset["Friuli-Venezia Giulia"]),
           np.mean(dataset["Piemonte"]),
           np.mean(dataset["Lombardia"]),
           np.mean(dataset["Trentino-Alto Adige"]),
            np.mean(dataset["Valle d'Aosta"]),
            np.mean(dataset["Liguria"]),
            np.mean(dataset["Emilia-Romagna"])]
# Manual division of macro-areas
f, p = stats.f_oneway(df_nord, df_sud, df_centro)
alpha = 0.05
alpha_half = alpha/2
print("ANOVA test on Organic Waste")
#print("Variance ratio:", max(np.var(df_nord), np.var(df_sud), np.var(df_centro))/min(np.var(df_nord), np.var(df_sud), np.var(df_centro)))
print("p value = {:g}".format(p))
print("f value = {:g}". format(f))
if p < alpha: # null hypothesis: the mean are the same
    print("The null hypothesis can be rejected")
else:
    print("The null hypothesis is accepted")
```

Listing A.21: T-tests on waste categories

A.5 Machine Learning

```
import pandas as pd
import pandas_datareader as pdr
import numpy as np
import matplotlib
import matplotlib.pyplot as plt
import seaborn as sns
import math
from statsmodels.tsa.api import acf, graphics, pacf
from statsmodels.tsa.ar_model import AutoReg, ar_select_order
from sklearn import metrics
from statsmodels.tsa.api import acf, graphics, pacf
from statsmodels.tsa.ar_model import AutoReg, ar_select_order
from sklearn import metrics
```

Listing A.22: Imports to perform Machine Learning

```
data = pd.read_excel('time series/96-20_rifiuti urbani.xlsx')
data.index = data["Unnamed: 0"]
data = data.drop("Unnamed: 0", axis = 1)
#Visualise Pre Regression Graphs
fig, ax = plt.subplots(figsize = (7, 4.5))
sns.despine()
ax.set_title('National solid waste production (1996-2020)')
ax.set_xlabel("years")
ax.set_ylabel("kg/per capita")
plt.grid(axis = "x")
sns.scatterplot(y = data.loc["Italia"][:] , x=data.columns, ax = ax, color = "purple")
# Polynomial linear regression

x = np.array(data.columns, dtype = float).reshape((-1, 1))
y = np.array(data.loc["Italia"][:])

degree = 2 #changing to 3/4

transformer = PolynomialFeatures(degree= degree, include_bias=True)
transformer.fit(x)
x_ = transformer.transform(x)
x_train,x_test,y_train,y_test= train_test_split(x_, y,test_size=0.2)

model = LinearRegression().fit(x_train, y_train)

print('Model coefficients are:', model.coef_)

y_pred = model.predict(x_test)
```

```

print('Mean Absolute Error, MAE is ', metrics.mean_absolute_error(y_test, y_pred))
print('Mean Squared Error, MSE is ', metrics.mean_squared_error(y_test, y_pred))
print('Root Mean Square Error is ', np.sqrt(metrics.mean_squared_error(y_test, y_pred)
    ))

#Rsquared value of model
print("Coefficient of determination or R squared value is ", metrics.r2_score(y_test,
    y_pred))
print("\n")

#New years
predicted_years = range(2021, 2026)
predicted_years = np.array(predicted_years).reshape((-1, 1))
predicted_years_ = PolynomialFeatures(degree=degree, include_bias=True).fit_transform(
    predicted_years)
y_pred = model.predict(predicted_years_)
#print('predicted response:', y_pred, sep='\n')
```

Listing A.23: Polynomial Regression

```
###Plotting
years = np.append(x, predicted_years)
prediction = np.append(y, y_pred)

df = pd.DataFrame({"rates": prediction, "years": years}, index = years)

#Visualise Polynomial Regression Graphs
fig, ax = plt.subplots(figsize = (7, 4.5))
ax.set_title('Estimation of national solid waste production (1996-2025)')
ax.set_xlabel("years")
ax.set_ylabel("kg/per capita")
plt.grid(axis = "x")
sns.despine()
sns.scatterplot(y = df["rates"], x = df["years"], ax = ax, color = "purple")
```

Listing A.24: Plotting the regression and prediction graph

```
data = pd.read_excel('time series/96-20 % diff.xlsx')
data.index = data["Unnamed: 0"]
data = data.drop("Unnamed: 0", axis = 1)

x = np.array(data.columns, dtype = float).reshape((-1, 1))
y = np.array(data.loc["Italia"][:])

x_train,x_test,y_train,y_test= train_test_split(x, y,test_size=0.2)

```

```
model = LinearRegression().fit(x_train, y_train)

print('Model coefficients are:', model.coef_)

y_pred = model.predict(x_test)

print('Mean Absolute Error, MAE is ', metrics.mean_absolute_error(y_test, y_pred))
print('Mean Squared Error, MSE is ', metrics.mean_squared_error(y_test, y_pred))
print('Root Mean Square Error is ', np.sqrt(metrics.mean_squared_error(y_test, y_pred)
    ))

#Rsquared value of model
print("Coefficient of determination or R squared value is ", metrics.r2_score(y_test,
    y_pred))
print("\n")

#New years
predicted_years = range(2021, 2026)
predicted_years = np.array(predicted_years).reshape((-1, 1))
y_pred = model.predict(predicted_years)

#Visualise Polynomial Regression Graphs
fig, ax = plt.subplots(figsize = (7, 4.5))
ax.set_title('Estimation of recycing rate Liguria (1996-2025)')
ax.set_xlabel("years")
ax.set_ylabel("kg/per capita")
plt.grid(axis = "x")
sns.despine()
sns.lmplot(data=df_ss, y = "rates", x = "years", order = degree)
```

Listing A.25: Linear Regression

A.6 Deep Learning

```
import numpy as np
import pandas as pd
import random
import os
import matplotlib.pyplot as plt
import seaborn as sns
import keras.applications.xception as xception
import zipfile
import sys
import time
```

```python
import tensorflow.keras as keras
import tensorflow as tf
import re

from PIL import Image
from keras.layers import Input, Conv2D, Dense, Flatten, MaxPooling2D, Input,
    GlobalAveragePooling2D
from keras.layers.experimental.preprocessing import Normalization
from keras.models import Model, Sequential
from keras.preprocessing import image
from keras.utils import to_categorical
from keras.layers import Lambda
from keras.callbacks import EarlyStopping
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report
```

Listing A.26: Imports to perform Deep Learning

```python
# Increasing the image size didn't result in increasing the training accuracy
IMAGE_WIDTH = 320
IMAGE_HEIGHT = 320
IMAGE_SIZE=(IMAGE_WIDTH, IMAGE_HEIGHT)
IMAGE_CHANNELS = 3
# Path where our data is located
base_path = "../input/garbage-classification/garbage_classification/"
# Dictionary to save our 12 classes
categories = {0: 'paper', 1: 'cardboard', 2: 'plastic', 3: 'metal', 4: 'trash', 5: '
    battery',
              6: 'shoes', 7: 'clothes', 8: 'green-glass', 9: 'brown-glass', 10: 'white
    -glass',
              11: 'biological'}
```

Listing A.27: Defining costants to perform DL algorithms

```python
# Add class name prefix to filename. So for example "/paper104.jpg" become "paper/
    paper104.jpg"
def add_class_name_prefix(df, col_name):
    df[col_name] = df[col_name].apply(lambda x: x[:re.search("\d",x).start()] + '/' +
    x)
    return df
# list conatining all the filenames in the dataset
filenames_list = []
# list to store the corresponding category, note that each folder of the dataset has
    one class of data
categories_list = []

for category in categories:
```

```
    filenames = os.listdir(base_path + categories[category])

    filenames_list = filenames_list  +filenames
    categories_list = categories_list + [category] * len(filenames)

df = pd.DataFrame({
    'filename': filenames_list,
    'category': categories_list
})

df = add_class_name_prefix(df, 'filename')

# Shuffle the dataframe
df = df.sample(frac=1).reset_index(drop=True)

print('number of elements = ' , len(df))
```

Listing A.28: Creating the DataFrame

```
df_visualization = df.copy()
df_visualization['category'] = df_visualization['category'].apply(lambda x:categories[
    x] )
df_visualization['category'].value_counts().plot.bar(x = 'count', y = 'category' )
plt.xlabel("Garbage Classes", labelpad=14)
plt.ylabel("Images Count", labelpad=14)
plt.title("Count of images per class", y=1.02);
```

Listing A.29: Categories distribution

```
early_stop = EarlyStopping(patience = 2, verbose = 1, monitor='
    val_categorical_accuracy' , mode='max', min_delta=0.001, restore_best_weights =
    True)
callbacks = [early_stop]
```

Listing A.30: Early stopping callback

```
from keras.models import Sequential
from keras.layers import Conv2D, MaxPooling2D, Dropout, Flatten, Dense, Activation,
    BatchNormalization
import keras.applications.densenet as DenseNet121

dense_layer = DenseNet121.DenseNet121(include_top = False, input_shape = (IMAGE_WIDTH,
     IMAGE_HEIGHT,IMAGE_CHANNELS),
                                   weights = "imagenet")
#Weights parameter defines the TransferLearning
# We don't want to train the imported weights
dense_layer.trainable = False
```

```
model = Sequential(name = "DenseNet")
model.add(keras.Input(shape=(IMAGE_WIDTH, IMAGE_HEIGHT, IMAGE_CHANNELS)))

#create a custom layer to apply the preprocessing
def dense_preprocessing(img):
    return xception.preprocess_input(img)

model.add(Lambda(dense_preprocessing))

model.add(dense_layer)
model.add(tf.keras.layers.GlobalAveragePooling2D())
model.add(Dense(len(categories), activation='softmax'))

model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['
    categorical_accuracy'])

model.summary()
```

Listing A.31: Densenet model

```
from keras.models import Sequential
from keras.layers import Conv2D, MaxPooling2D, Dropout, Flatten, Dense, Activation,
    BatchNormalization
import keras.applications.mobilenet as MobileNet

dense_layer = MobileNet.MobileNet(include_top = False, input_shape = (IMAGE_WIDTH,
    IMAGE_HEIGHT,IMAGE_CHANNELS),
                                   weights = "imagenet")
#Weights parameter defines the TransferLearning
# We don't want to train the imported weights
dense_layer.trainable = False
model = Sequential(name = "MobileNet")
model.add(keras.Input(shape=(IMAGE_WIDTH, IMAGE_HEIGHT, IMAGE_CHANNELS)))

#create a custom layer to apply the preprocessing
def dense_preprocessing(img):
    return xception.preprocess_input(img)

model.add(Lambda(dense_preprocessing))

model.add(dense_layer)
model.add(tf.keras.layers.GlobalAveragePooling2D())
model.add(Dense(len(categories), activation='softmax'))

model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['
    categorical_accuracy'])

```

```
model.summary()
```

Listing A.32: Mobilenet model

```
#Change the categories from numbers to names
df["category"] = df["category"].replace(categories)

# We first split the data into two sets and then split the validate_df to two sets
train_df, validate_df = train_test_split(df, test_size=0.2, random_state=42)
validate_df, test_df = train_test_split(validate_df, test_size=0.5, random_state=42)

train_df = train_df.reset_index(drop=True)
validate_df = validate_df.reset_index(drop=True)
test_df = test_df.reset_index(drop=True)

total_train = train_df.shape[0]
total_validate = validate_df.shape[0]

print('train size = ', total_validate , 'validate size = ', total_validate, 'test size
    = ', test_df.shape[0])
```

Listing A.33: Test/train set

```
batch_size=128          #32,64,128,256
train_datagen = image.ImageDataGenerator()
train_generator = train_datagen.flow_from_dataframe(
    train_df,
    base_path,
    x_col='filename',
    y_col='category',
    target_size=IMAGE_SIZE,
    class_mode='categorical',
    batch_size=batch_size
)
validation_datagen = image.ImageDataGenerator()
validation_generator = validation_datagen.flow_from_dataframe(
    validate_df,
    base_path,
    x_col='filename',
    y_col='category',
    target_size=IMAGE_SIZE,
    class_mode='categorical',
    batch_size=batch_size
)
```

Listing A.34: Test/train set

```
EPOCHS = 20
history = model.fit_generator(
    train_generator,
    epochs=EPOCHS,
    validation_data=validation_generator,
    validation_steps=total_validate//batch_size,
    steps_per_epoch=total_train//batch_size,
    callbacks=callbacks
)
```

Listing A.35: Training the model

```
import seaborn as sns

sns.set(rc={'figure.figsize':(7.2,6)})
sns.set(style="whitegrid")

fig, (ax1, ax2) = plt.subplots(2, 1)
ax1.plot(history.history['loss'], label="Training loss", color = "olivedrab")
ax1.plot(history.history['val_loss'], label="validation loss", color = "rebeccapurple"
    )
ax1.set_yticks(np.arange(0, 0.7, 0.1))
ax1.grid(axis = "x")
ax1.legend()

ax2.plot(history.history['categorical_accuracy'], label="Training accuracy", color = "
    olivedrab")
ax2.plot(history.history['val_categorical_accuracy'], label="Validation accuracy",
    color = "rebeccapurple")
ax2.legend()
ax2.grid(axis = "x")
legend = plt.legend(loc='best')
plt.tight_layout()
sns.despine()
plt.show()
```

Listing A.36: Plotting results