**Capstone Project**

**Document Skeleton**

# Process overview

The following diagram shows the overall end-to-end process for defining, designing and delivering the Capstone project.



# Problem statement

* What is the problem or the opportunity that the project is investigating?

In 2018-19 Australia generated an estimated 74 million tonnes of waste i.e. 2.94 tonnes per person and only 60% was recycled. 40% found its way to landfills which receive around 20 million tonnes of waste each year.

Glass, plastic, paper, cardboard, tyres, electronics, batteries, construction materials, food and garden waste are all waste items containing value in the form of energy and resources.

Yet the amount of rubbish we create is constantly increasing because:

* Increasing wealth means that people are buying more products and ultimately creating more waste.
* Increasing population means that there are more people on the planet to create waste.
* New packaging and technological products are being developed, much of these products contain materials that are not biodegradable.
* New lifestyle changes, such as eating fast food, means that we create additional waste that isn’t biodegradable.
* Why is this problem valuable to address?

#### **Environmental Importance**

waste has a huge negative impact on the natural environment.

* Harmful chemicals and greenhouse gasses are released from rubbish in landfill sites. Recycling helps to reduce the pollution caused by waste.
* **Recycling** reduces the **need** for extracting, refining and processing raw materials all of which create air and water pollution while destroying natural habitats e.g. rainforests
* Huge amounts of energy are used when making products from raw materials. Recycling requires much less energy and therefore helps to preserve natural resources. As **recycling** saves energy it also reduces greenhouse gas emissions, which helps to tackle climate change

**Importance To People**

Recycling is essential to cities around the world and to the people living in them.

* No space for waste. Landfills are getting bigger and filling up faster in a world where land is becoming scarce due to unprecedented population growth rates.
* Preserve natural resources for future generations. Recycling reduces the need for raw materials; it also uses less energy, therefore preserving natural resources for the future.

**Economic Value**

* Reduce financial expenditure in the economy. Making products from raw materials costs much more than if they were made from recycled products.
* What is the current state (e.g. unsatisfied customers, lost revenue)?

|  |  |
| --- | --- |
|  | * 1. Linear flow of materials – ‘take, make, use and dispose’ model.   2. Limited use of renewable materials and energy   3. Significant volumes of materials disposed of and lost to the economy.   4. Limited focus on life cycle thinking. |

* What is the desired state?

|  |  |
| --- | --- |
|  | 1. Circular flow of materials - – materials sorted and retained in the economy for as long as possible. 2. Preference for renewable materials and energy. 3. Materials recovered as high up the waste hierarchy as possible. 4. Products designed and manufactured to minimise environmental impact through whole of life. |

* Has this problem been addressed by other research projects? What were the outcomes?

Waste management and recovery problem is as old as modern human settlements. Numerous research and approaches have gradually contributed to solving this behemoth of a crisis.

Recent years the advent of computer vision and DL has seen a renewed approach to the problem. The application of CV and AI in this sector is still in its infancy.

# Industry/ domain

* What is the industry/ domain?

Environment and natural resources

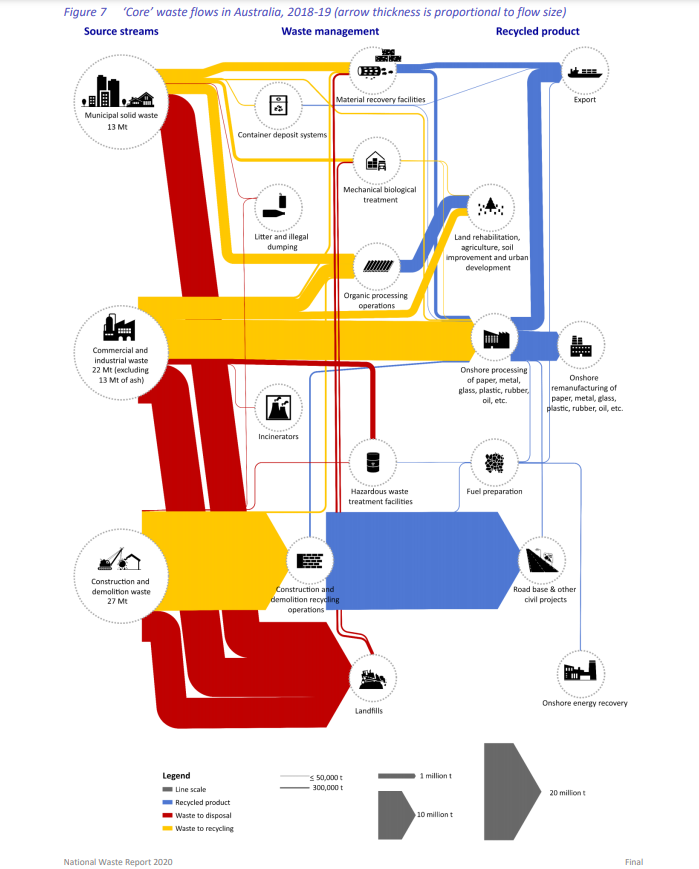
* What is the current state of this industry? (e.g. challenges from startups)

Increasing amounts of waste is dumped in landfills each year causing damage to environment.

Low recycles rates of waste

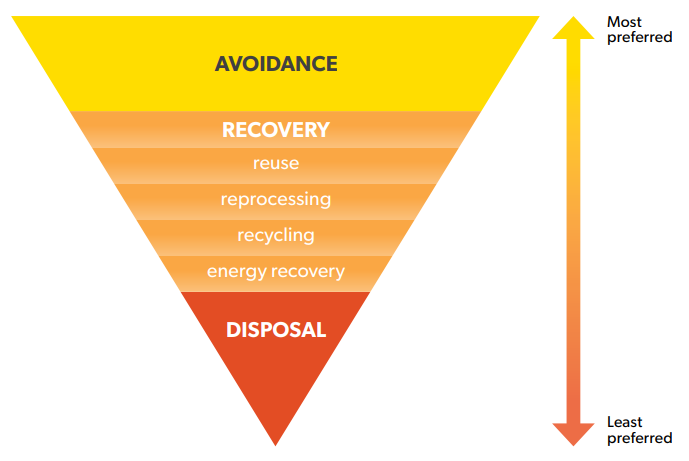
Resource intensive economy

* What is the overall industry value-chain?



* What are the key concepts in the industry?

The waste hierarchy ranks waste management options in order of their general environmental desirability



Waste avoidance is the most preferred option in the hierarchy.

Resource recovery options recover value from materials, thereby offsetting the environmental impacts of extracting and processing raw materials. Energy recovery is the least preferred recovery option.

Disposal is the least preferred option. Disposal generally recovers the least value from materials and delivers the least environmental benefit

* Is the project relevant to other industries?

Yes, to any industry and facet of society where materials are consumed, and solid waste generated.

# Stakeholders

* Who are the stakeholders? (be as specific as possible)

Department of Agriculture, Water and the Environment

Local government and councils

Material recovery facilities

Environmentalists

* Why do they care about this problem?

Waste disposal into landfills has negative environmental impact conflicting with environmental and social responsibility. It is increasingly becoming an ethics issue.

Inability to reuse recyclable material has a significant economic impact.

* What are the stakeholders’ expectations?

Environmentalists – zero waste, carbon neutral economies

Recyclers – high quality waste material i.e. no contamination

Material handling facilities – improved waste sorting/separation efficiency, reduced operating costs

Government departments, councils – increased compliance to waste management policies

# Business question

* What is the main business question that needs to be answered?

How can we reduce the amount of recyclable material which end up in landfills.

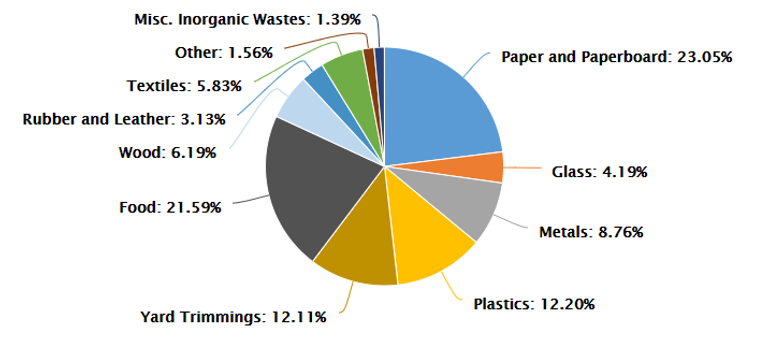
* What is the business value of answering this question? (quantify value and make necessary assumptions)

Reduced environmental impact and greenhouse emissions – priceless

Recover and recycle useful materials. A 25% recovery of recyclable material dumped equates to a retained value of 300million per year to economy.

***Assumptions***

Typical landfill waste composition from USA was obtained. Assuming that this is similar or closely related to that of Australia and using nominal recycle commodity process and a soft 25% recovery target, the economic value was thus calculated as below.





<https://www.afr.com/companies/infrastructure/recycling-bad-habits-costing-australia-324-million-20190910-p52pyz>

<https://ksenvironmental.com.au/recycling-commodity-prices-the-china-effect/>

* What is the required accuracy? What are the implications of false positives or false negatives?

roc\_auc\_score is the accuracy of interest

False positives/negative will only result in reduced recycle quality (contamination). This is not a catastrophic issue.

# Data question

* What is the data question that needs to be answered?

Build an image classification model.

* What is the data required to answer the question?

Labelled images of the different waste classes

# Data

* Where was the data sourced?

Trashnet dataset by Gary Ching and Mindy Yang, Stanford University

<https://github.com/garythung/trashnet/tree/master/data>

The dataset is also found on kaggle

<https://www.kaggle.com/asdasdasasdas/garbage-classification>

* What is the volume and attributes of the data?

images

6 classifications: cardboard (393), glass (491), metal (400), paper(584), plastic (472) and trash(127

* How reliable is the data?

The data is reliable but needs to be expanded to include deformed and or contaminated material to train a more robust and production ready model.

* What is the quality of the raw data?

Images, clean, slight imbalance

* How was this data generated?

Photographs / Images captured of materials and labelled

Image augmentation applied to generate more images and increase training dataset

* Is this data available on an ongoing basis?

Yes, it is available.

# Data science process

## Data analysis

* What data pipeline was to wrangle the raw data?

1. **Read image**

* CV2 to read image files and convert to right colour format (BGR to RBG)
* Resize image into input dimension for models i.e. 150x150x3

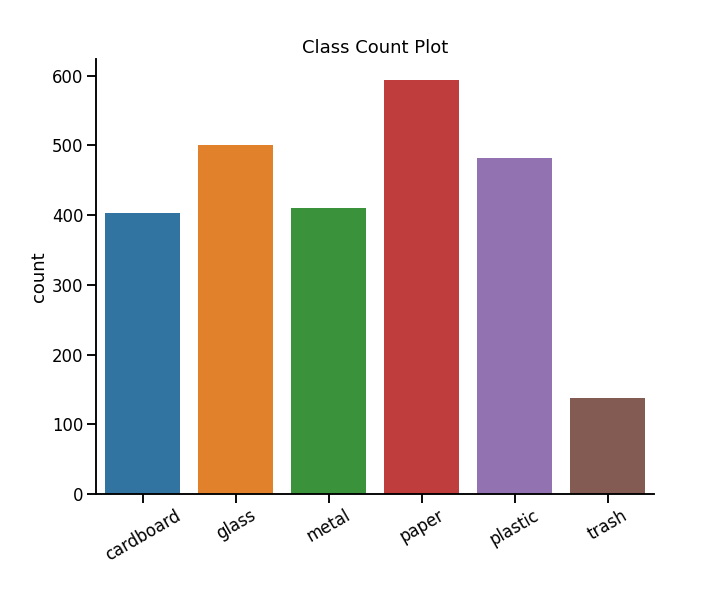
1. **Format**

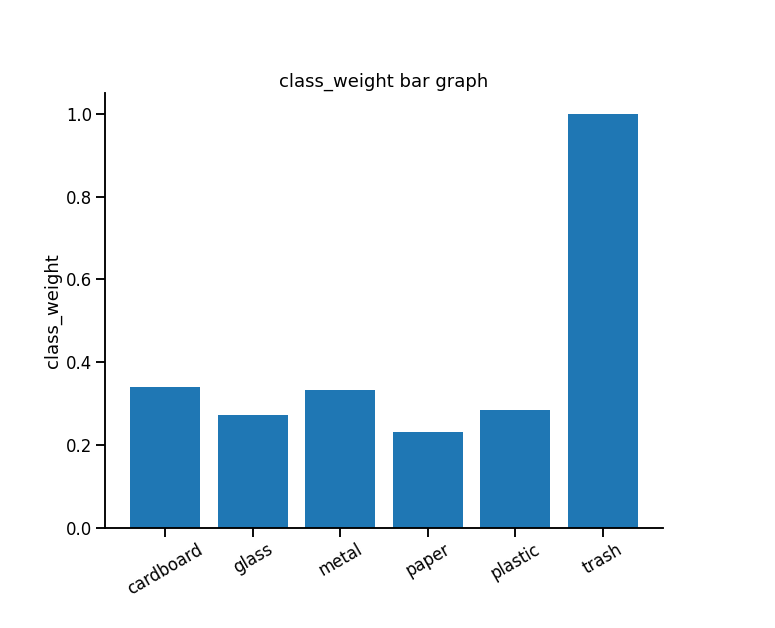
* Covert image into NumPy array of datatype ‘float32’
* Normalize image array values for faster convergence
* Create array of labels and label- encode --🡪 one-hot-encode

**3. Split data**

* train\_validate\_test split
* ImageDataGenerator: image augmentation (rotate,width\_shift,height\_shift, hosizontal\_flip, vertical\_flip)
* What are the highlights of the Exploratory Data Analysis (EDA)?

Slightly imbalanced dataset





Inverse Number of Samples to determine class\_weights to address dataset imbalance.

* Is the pipeline reusable? (for example, to process future data?)

Yes, it is. Images need to be labelled are required to be saved in separate folders

* What are the intermediary data structures used (if any)?

N/A

## Modelling

* What are the main features used?

N/A

image classification

Images resized to - 150 x 150 x 3

* Did you find any interesting interactions between features?

N/A

* Is there a subset of features that would get a significant portion of your final performance? Which features?

N/A

* How did you select features?

N/A

* What feature engineering techniques are used?

N/A

* What are the models used?

1. CNN model - 18 layered
2. VGG16
3. ResNet

* How long does it take to train your model?
* What are the tools used? (Cloud platform, for example)

Jupyter Lab, GoogleColab,

* What are the model performance metrics?

Accuracy\_score

roc\_auc\_score

* Which model was selected?

Vgg16 model with transferred learning

## Outcomes

* What are the main findings and conclusions of the data science process?

Vgg16 with transferred learning models can be successfully trained for image classification of waste.

## Implementation

* What are the considerations for implementing the model in production?



# Data answer

* Was the data question answered satisfactorily?
* What is the confidence level in the data answer?

# Business answer

* Was the business question answered satisfactorily?
* What is the confidence level in the business answer?

# Response to stakeholders

* What are the overall messages and recommendations to the stakeholders?

# End-to-end solution

* What is the overall end-to-end solution to use the model developed in the project?

# References

* Where are the data and code used in the project? (show a simplified list of main items: notebooks, datasets, exported models)
* What are the resources used in the project? (libraries, algorithms, etc)