**Identifying business goals**

As a normal human being, it can be so hard to sort rubbish. You sometimes can’t tell if you can recycle this particular item or not. Personally, it drives me crazy when this happens! Should I throw it in the rubbish bin or not? Should I take my chance and recycle it? I never quite know most of the time. Can I even do anything about it?

It is not only me, though, as machinery is not exempted from this. They also have trouble sorting out if an item is recyclable or not due to common items nowadays comprising of different materials.

The questions to be asked is this: how should we effectively separate recycling?

What seems to be the major problem in the recycling separation process? Well, it really falls from the fact that it is difficult to sort out the rubbish. When anything takes a considerable effort, humans tend to find a way to not do it!

Thankfully, technology is rapidly advancing that even serious problems like pollution are bound to be solved by the rapidly advancing technology created nowadays. With the help of technology, we can make improvements!

A way to revolutionize the recycling separation process is surely by [utilizing the technology](https://www.bbc.com/news/av/business-50335737/could-invisible-barcodes-revolutionise-recycling) that we have today.

**Business Understanding**

*Background*

One of the most important steps of waste management is the separation of the waste into the different components and this process is normally done manually by handpicking. Typically, recycling facilities have an implemented process to carefully separate recyclables from the non-recyclable ones, sorted further and then recycle.

First, they will collect all the waste and take them to Recover Facility. There, they then open the bags that contain the waste and put them onto a conveyor belt where they start the separating process. They sort out the materials – they inspect the items and determine if they can recycle it or not. Those non-recyclable are removed.

Every recyclable materials are further sorted and processed differently from other recyclables. Afterwards, the materials make their way to different industries.

*Business Objectives*

Our goal is to feed a database and create a machine-learning model in order to help the correct classification of waste and recycling, and thus create conscience that the new technology that is already being implemented in Europe of an invisible barcode printed in the packages will help us improve even more our capacity of recycling

*Success Criteria*

Feed the database of a test recycling plant in order to improve the self-adapting sequential sorting system that sorts the waste in loops according to the predominant material thus allow the operators to correctly classify 90% of the incoming materials for recycling and reduce human involvement.

**Assess the Current Situation**

*Inventory of Resources*

Waste image samples previously compiled during the implementation process of a tender project for development and design of a waste treatment plant

Images collected and a repository of datasets for review

<https://github.com/garythung/trashnet>

<https://www.bbc.com/news/av/business-50335737>

<https://youtu.be/g_ajkE77Nik>

*Requirements, assumptions and constraints*

Before building, training and evaluating our model, we must gather a bigger dataset of labeled waste images. Based on the original data I already have, a github repository full of datasets about waste and photographies that I can take with a few web tools that would help with the search and then save and rename the files with windows tools…

*Risks and contingencies*

As we are dealing with a relatively small dataset, when training a model we’d need to process a pre-trained model or apply other algorithms that would help us to artificially introduce sample diversity by applying random, yet realistic, transformations to the training image

In addition, due that part of our dataset would be from Google, we could find the problem that many of the images would come with watermark or other kind of mark that would make them useless to use or unrecognized by TF.

*Terminology*

**Aluminum:** cans and other packaging, furniture, appliances

**Glass:** glass bottles, glass jars, tops removed

**Organic waste**: food waste, fruit peelings, waste from vegetables and root crops, eggshells, coffee grounds and tea leaves, along with filter papers, waste from fish and small bones, solid fats, kitchen roll and paper napkins, flower soil and plant waste, wood chips and sawdust (not from impregnated wood), wood-based pet bedding, paper bags or newspaper used as a drying agent

**Other Plastics**: nondurable products, such as disposable diapers, trash bags, cups, utensils, medical devices and household items such as shower curtains. The plastic food service items are generally made of clear or foamed polystyrene, while trash bags are made of high-density polyethylene (HDPE) or low-density polyethylene (LDPE).

**Paper and Cardboard:** newspapers and magazines, advertisements, brochures, envelopes (including those with a window), photocopy papers and printouts (coloured and white), drawing paper and paper from pads (coloured and white), recycled paper, phone directories and product catalogues, and paperback books and pages from hardback books (with covers removed), notebooks, also with metal or plastic bindings, paper bags made of white paper.

Carton material should be empty, clean, dry, flattened and placed inside one box, corrugated cardboard, cardboard containers for ready-made meals, cardboard containers for dry products, including cereal and biscuit boxes, paper bags, pizza boxes and egg boxes, disposable cardboard plates and cups, toilet paper and kitchen roll tubes, wrapping paper, such as for photocopy paper, cardboard sleeves from drinks multipacks

**Plastic:** This category includes bags, sacks and wraps; other packaging; polyethylene terephthalate (PET) bottles and jars; high-density polyethylene (HDPE) natural bottles; and other containers. Manufacturers also use plastic in durable goods, such as appliances, furniture, casings of lead-acid batteries and other products.

**Textiles**: material that becomes unusable or worthless after the end of the production process of any textile product. Wastage produce in every stage of the textile manufacturing process such as [spinning](https://textileapex.blogspot.com/2014/01/ring-spinning-definition-object.html), [weaving](https://textileapex.blogspot.com/2014/08/weaving-woven-fabric-history.html), [knitting](https://theswatchbook.offsetwarehouse.com/2017/05/04/knitting-brief-history-knitting-uses/), [dyeing](https://textileapex.blogspot.com/2014/03/dyeing.html), [finishing](https://textileapex.blogspot.com/2014/09/garment-finishing.html) and [clothing](https://en.wikipedia.org/wiki/Clothing).

**Wood:** The sources of wood in municipal solid waste (MSW) include furniture, other [durable goods](https://www.epa.gov/facts-and-figures-about-materials-waste-and-recycling/durable-goods-product-specific-data) (e.g., cabinets for electronic equipment), wood packaging (crates, pallets) and some other miscellaneous products.

**Determine Data Mining Goals**

*Data Mining Goals*

Dataset size of at least 250 labeled images of the each of waste

Desired dataset size of 500 labeled images of each type of waste

*Data-mining success criteria*

We don't have a large image dataset, it’d help to artificially introduce sample diversity by applying random, yet realistic, transformations to the training images, such as rotation and horizontal flipping. This helps expose the model to different aspects of the training data and reduce [overfitting](https://www.tensorflow.org/tutorials/keras/overfit_and_underfit).

We’ll need enough images labelled and organized for each type of waste in format .jpg, preprocessed and resized for better computing

If there’s images not recognized by TF, we’ll implement a line of code that help us identifying this images

**Produce Project Plan**

*Project Plan*

Organize and label images for each type of waste.

Explore, organize and rename new images according with the original labels

After having the data ready, build the Colab Notebook with all the Python code

Apply a line of code that will help us identify if there’s files that can’t be used y TF

Split the data into training and validation datasets

Autotune the datasets using buffered prefetching to load images from disk without having I/O become blocking.

Use data augmentation, this helps expose the model to different aspects of the training data and reduce [overfitting](https://www.tensorflow.org/tutorials/keras/overfit_and_underfit).

Pre-trained models expects pixel values in [-1, 1], but at this point, the pixel values in our images are in [0, 255]. To rescale them, use the preprocessing method included with the model.

Create the base model from a pre-trained convolutional network

Compile and Train model

Fine tune model

Model evaluation

*Initial Assessment of Tools and Techniques*

Google Colab, Jupiter Notebook, Chrome extensions, powershell