# Abstract [*≈* 1 paragraph] (Luigi)

The abstract is optional, depending on your available space. It should consist of 0.5 paragraph consisting of the motivation for your paper and a high-level explanation of the methodology you used/results obtained.

1. **Introduction [***≈* 0*.*5 **pages] (Luigi)**

Explain the problem and why it is important. Discuss your motivation for pursuing this problem. Give some background if necessary. Clearly state what the input and output is. Be very explicit: “The input to our algorithm is an *{*image, amplitude, patient age, rainfall measurements, grayscale video, etc.*}*. We then use a *{*SVM, neural network, logistic regression, etc.*}* to output a predicted *{*age, stock price, cancer type, music genre, etc.*}*.” This is very important since different teams have different inputs/outputs spanning different application domains. Being explicit about this makes it easier for readers. If you are using your project for multiple classes, add a paragraph explaining which components of the project were used for each class.

The production of cars has been steadily increasing in the past decade and this has given rise to the used car market. The emergence of online second-hand car portals in the United Kingdom and Europe has facilitated the need for both the customer and the seller to be better informed about trends and patterns that determine the value of a used car [1]. A second-hand car price prediction system (for a specific car brand) is thus required to effectively determine the worthiness of the car using a variety of features. The price prediction model and insights/patterns could be then later used as a tool for a second-hand car retailer to give insights to potential customers shopping for a second-hand car.

To tackle the need of a price prediction system, we are going to divide the price prediction problem into two subproblems: a price prediction model for a specific car brand (Mercedes C-Class W205) and a price prediction model for an entire brand (Mercedes). Our main goal is thus to make a price prediction model both a specific model and an entire brand and extract some interesting insights of the entire dataset (all cars and models combined).

For our first price prediction model of the Mercedes C-Class W205, we predict the price (output) based on the mileage (input). We first implement regularized linear regression and will further improve our model by introducing polynomial regression. For our second price prediction model of Mercedes cars, we predict the price (output) based on several features (year, fuel type, transmission type, etc.) by using multivariate regression.

1. **Related work [***≈* 0*.*5 **pages] (Luigi)**

You should find existing papers, group them into categories based on their approaches, and discuss their strengths and weaknesses, as well as how they are similar to and differ from your work. In your opinion, which approaches were clever/good? What is the state- of-the-art? Do most people perform the task by hand? You should aim to have at least 5 references in the related work. Include previous attempts by others at your problem, previous technical methods, or previous learning algorithms. Google Scholar is very useful for this: <https://scholar.google.com/>(you can click “cite” and it generates MLA, APA, BibTeX, etc.)

1. **Dataset and Features [***≈* 0*.*5 *−* 1 **pages] (Luigi)**

Describe your dataset: how many training/validation/test examples do you have? Is there any preprocessing you did? What about normalization or standardization? What is the resolution of your images? How is your time-series data discretized? Include a citation on where you obtained your dataset from. Depending on available space, show some examples from your dataset. You should also talk about the features you used. If you extracted features using Fourier transforms, histogram of oriented gradients (HOG), PCA, ICA, etc. make sure to talk about it. Try to include examples of your data in the report (e.g. include an image, show a waveform, etc.).

1. **Methods [***≈* 1 *−* 1*.*5 **pages] (Kai)**

Describe your learning algorithms, proposed algorithm(s), or theoretical proof(s). Make sure to include relevant mathematical notation. For example, you can briefly include the SVM optimization objective/formula or say what the softmax function is. It is okay to use formulas from the lecture notes. For each algorithm, give a short description (*≈* 1 para- graph) of how it works. Again, we are looking for your understanding of how these machine learning algorithms work. Although the teaching staff probably know the algorithms, future readers may not (reports will be posted on the class website). Additionally, if you are using a niche or cutting-edge algorithm (e.g. long short-term memory, SURF features, or anything else not covered in the class), you may want to explain your algorithm using 1/2 paragraphs.

# Results/Discussion [*≈* 1 *−* 3 pages] (Kai)

You should also give details about what (hyper)parameters you chose (e.g. why did you use X learning rate for gradient descent, what was your mini-batch size and why) and how you chose them. Did you do cross-validation, if so, how many folds? Before you list your results, make sure to list and explain what your primary metrics are: accuracy, precision, etc. Provide equations for the metrics if necessary. For results, you want to have a mixture of tables and plots. You should include a confusion matrix or AUC/AUPRC curves. Include performance metrics such as precision, recall, and accuracy. Include visualizations of results, heatmaps, examples of where your algorithm failed and a discussion of why certain algorithms failed or succeeded. In addition, explain whether you think you have overfit to your training set and what, if anything, you did to mitigate that. Make sure to discuss the figures/tables in your main text throughout this section. Your plots should include legends, axis labels, and have font sizes that are legible when printed.

# Conclusion/Future Work [*≈* 1 *−* 2 paragraphs] (Luigi)

Summarize your report and reiterate key points. Which algorithms were the highest- performing? Why do you think that some algorithms worked better than others? For future work, if you had more time, more team members, or more computational resources, what would you explore?

All sections before this point must fit on six (6) pages. No exceptions. Supplemental material is not allowed. Anything else you want to add to your report (e.g. acknowledgements, author bios, funding sources) is included in the 6 page limit. The report should be in **1 column**. The exception is the section describing the contributions of each team member. **You will be penalized** *−*11 **points per page exceeding this limit**. The max report score is 100.

# Contributions

The contributions section is not included in the 6 page limit. This section should describe what each team member worked on and contributed to the project.