The Models of London Bike Sharing Prediction

This manuscript ([permalink]) was automatically generated from [Jingzi2020/CEE498_2020_G1@2a67cba] on December 6, 2020.

Authors

- **Jingzi Chen**
  [86-13522264917] · [Jingzi2020]
  Grainger College of Engineering, University of Illinois at Urbana-Champaign

- **Anye Wang**
  [15409980880] · [wanganye123]
  Grainger College of Engineering, University of Illinois at Urbana-Champaign, Civil Engineering - Construction Management

- **Dana Monzer**
  [+1 2178191851] · [Dana2021]
  Grainger College of Engineering, University of Illinois at Urbana-Champaign, Civil Engineering - Transportation Systems

Abstract

The aim of this study is to create a predictive model for bike-sharing counts in a n hour in the city of London in United Kingdom. The model makes use of regular neural network. And the main features affecting the bike counts include weather conditions and time. The model root mean square is 210, with a mean of 1124 counts in an hour compared to 1138 counts in an hour in the training data. The model provides enough accuracy for planning a new station and scheduling bike redistribution schedules.

Introduction

Bike-sharing system is an important mode in sustainable transportation modes. Great attention is put to improve such systems to increase their demand and maximize their environmental benefits along with other societal benefits. To study the demand of this system, London Bike-sharing system sis explored. The purpose of this study is to predict ranges of new bike counts each hour based on certain factors provided in the dataset. The aim of this report is to create a predictive model using machine learning to predict bike counts in a given hour in London city which is useful for planning and operation forecasting of bike-sharing system. The exploratory data analysis using graphical and statistical tools in Python were used to derive preliminary conclusions about the dataset by analyzing the results of the tools used. The dataset provided was acquired from three sources, to include the new bike counts in each hour, the weather conditions, and the holidays. The data from cycling dataset is grouped by "start time", and it represents the count of new bike shares grouped by hour. The long duration shares are not taken in the count." The data sample analyzed in this project is collected between January 1st, 2015 to January 1st, 2017 in London, UK and it includes the following parameters: - Timestamp (year, month, day, hour) - Cnt: the count of a new bike shares - T1: temperature measure taken in degree Celsius - T2: temperature feels - Hum: humidity percentage - Wind_speed: in Km/hr
- Weather_code: 1 = Clear ; mostly clear but have some values with haze/fog/patches of fog/ fog in vicinity 2 = scattered clouds / few clouds 3 = Broken clouds 4 = Cloudy 7 = Rain/ light Rain shower/ Light rain 10 = rain with thunderstorm 26 = snowfall 94 = Freezing Fog - Is_holiday: 1 if it is a holiday, 0 if it is not. - Is_weekend: 1 if it is a weekend, 0 if it is not.
- Season: 0: Spring, 1: Summer, 2: Fall, 3: Winter The sections below include literature review, description of methods, results, discussion and conclusion.

Literature Review

Bike sharing system has appeared more and more on the street of the cities in order to meet the demand of public transportation in the last short distance to the destination. Also, Bike sharing system is so popular around all the world that most of major modern cities and campuses have been operated. Among researches for the bike system, predicting the demand of future bike shares is one of the most important and necessary tasks to ensure a satisfactory level of service of the system.

Many recent studies have focused on the problem of predicting demand of shared bikes in the system. In order to complete prediction, the bike sharing system is supposed to satisfy the real-time and high accuracy requirements. Some researchers utilized Spark Machine Learning framework to predict the number of rental bikes to optimize the accuracy of model [1]. Firstly, the author collected three types of data including Citi Bike, Weather and Holiday, which the data have been SQL and outlier processed. Then, three predicative models including multiple linear regression, decision tree and random forests are constructed to analyze and train the processed data. During the experiment, applying to machine learning Spark ecosystem, the author used 70 % data as training data while 30%

as the test data. In the result, after testing, random forests model has the lowest root mean error (RMSE). At last, the researcher optimized the result further by applying logarithmic optimization to the model.

Throughout the literatures, there is consistent evidence that forecast number of shared bikes are influenced not by holiday, weather, and temperature [2] in general, but by user information such as gender, birth year and user type [1]. The strengths of the literature are that the author consider more factors like user information and process useful and reasonable row data. Moreover, the optimization improves the accuracy the result. While there has been much research on algorithm optimization, few researchers have taken the influence of different factors and the real-time data into consideration [2]. The weakness of the literature is that the author ignored the timeliness of data and station selection. Furthermore, the mythology is limited to predict the number in the new area where lack of the existing shared bike number. Usually, station clustering and demand prediction in every station should be evaluated and analyzed as an integration [3].

In general, it is challenging to collect the external factors in the future study because there are too many factors affecting users whether to use shared bike. However, we should consider multiple factors as much as possible to improve the accuracy. Comparing with our project's database, more factors are taken into consideration including time, number of new bike shares grouped by hour, temperature, humidity, wind speed, types of the weather, holiday, weekend, season [4]. In our project, we will employ historical usage of bikes with some necessary factors to predict the number, which are very similar with the literature because it is easier to collect data and the predicted numbers are closer to the real.

The second paper studied is “Case Studies on Transport Policy Modeling bike counts in a bike-sharing system considering the effect of weather conditions” [5]. This paper identifies a method to quantify the effect of weather conditions on bike sharing counts in San Francisco Bay area with the aim of improving bike sharing systems given their wide benefits such as decreasing transportation pollution and increasing mobility efficiency in cities, to name a few. The benefit of this model specifically will be to decrease the environmental costs and time consumption and other complications associated with the rebalancing operation of bikes between stations; which is important to ensure that each station has enough number of bikes to satisfy the demand, especially given the limited number of docks at station.

The methodology used to create the bike count model includes several steps. First the effect of various variables is quantified (month of the year, day of the week, time of the day and different weather conditions), then these predictors were ranked by Random Forest technique and were used to predict a regression model using a guided forwarded step-wise regression. More than one model was created then the Bayesian information criterion was used to evaluate the models. In the first step, the count models employed generalized linear models, specifically two models were used Poisson, which condition to apply is that the mean and variance must be equal, and negative binomial regression, which uses same condition as Poisson except that another parameter is involved that loosens the initial condition and adjusts the variance independently, so it is considered to accommodate more dispersion. The second step after that is to apply machine learning to avoid overfitting of predictors, using Random Forrest method. The RF method randomly constructs a group of trees, where each tree is a subset of features, so trees are not correlated, then the ranking of features is obtained based on majority of votes from all trees, after that forward step-wise regression is applied, and finally a model is selected BIC after computing the log-likelihood of each model, and the model of the lowest BIC is to be selected.

This methodology [5] was applied on a dataset of bike-sharing for San Francisco Bay Area between August 2013 to August 2015, where incidents were documented every minute for 70 stations in the area, which led to a large dataset, and another dataset was used which included weather conditions during these 2 years, and it included the following attributes: “date (in month/day/year format), ZIP code, temperature, humidity, dew level, sea level pressure, visibility, wind speed and direction, precipitation, cloud cover, and weather description for that day (i.e., rainy, foggy or sunny).”

For the first step, the histogram of new counts frequency for all stations showed dispersion, which gave a hint on better fitness of NBRM. However, both PRM and NBRM were applied at first to generate a full model of all available predictors. Then RF was used to rank the predictors in the full model based on the OOB error. Forward stepwise regression was then used to fit several models that were constructed by RF, then BIC was applied to select the best subset of predictors to construct this model. At first it was assumed that there is no interaction between the 70 stations, to quantify effects fast and effectively and in an attempt to create one model for all variables rather than a model for each station, and this approach was described to satisfy the level of accuracy needed. And the results showed a logarithmic mean of bike counts at each station following parallel hyperplanes, which shows no interaction between stations. And in order to construct one model instead of 70 models, one for each station: 69 indicators were used with one reference, a similar approach was used for months of year with 11 indicators and January as a reference, and 6 indicators for the days of the week, and so on for all data attributes. If there was no significant difference between each pair of parameters, for example between 2 stations, it was assumed bike count was the same for the two stations to an acceptable level of accuracy. The results showed that different stations, month-of-the-year, day-of-the-week, and time-of-the-day were all shown to influence the model. And the following weather attributes were selected for additional exploration: mean temperature, mean humidity, mean visibility, mean wind speed, precipitation, and weather description. And for the second time, RF and forward step-wise regression were employed and the resulting models were compared by BIC. And the model with the trade-off between the minimum BIC value and the consideration of the effective parameters was selected. And as was shown because of dispersion, indeed NBRM was shown to be better than PRM, and it was selected for the rest of the modeling steps. Among 111 models created, the results showed that bike counts are significantly influenced by the month-of-the-year, day-of-the-week, time-of-the-day, and some weather variables, mainly temperature and humidity level, which is also dependent on geographic location. And the most significant variables affecting bike counts are available number of bikes at time t-1 and the time-of-the-day.

This paper holds many strengths in meeting our project, first the factors used in this paper match our dataset, and the paper justifies the use of these factors among others used in other studies. And this paper was the first to study the effect of humidity which was shown to have a significant effect on the model. The methodology and tools used match our set of expertise, which is to be developed through this course, and the tools and methods used were all explained and justified. This paper can be used as a reference for us to build our model and compare the results given different factors, especially difference in geographic location, while also contributing to this research with the insights we obtain.

However, a few weaknesses of this paper were detected. First, the paper targets only docked bike-sharing model and it’s applicable to only certain geographic areas with certain weather conditions, as for another location, different parameters might be additionally considered. The final stages of refining the model and detecting possible errors was merely systematic and rather relied on observations and experts opinions, so in addition to the scientific and sequenced steps, there was some subjectivity in the methodology when it comes to certain decisions like the number of trees in RF and the selected factors for each tree and their number which wasn’t discussed and explained enough, so it might be hard for us to follow the same methodology at these stages. The paper only used Poisson and negative binomial models at the first stage and didn’t attempt more complex distributions. And the paper chose the final model prioritizing simplicity, designated by a smaller number of predictors, among the last two proposed models, without explicitly comparing their levels of accuracy. And some standard values of comparisons for example minimum log-likelihood and typical BIC measures weren’t shared in the paper.

Another study published by Lin, Wang, Jiang, Fan and Sun (2017) [6], tried to figure out sharing bike demand's prediction based on the Bayesian classifier and APSO-BP neural network models. In this article, the researchers collected weather data and historic public bicycle sharing record to build the dataset. Then, the model categorize bicycle rental mode through Bayesian classifier and used the specific neural network fitted to the mode. The evaluation of this model was based on its accuracy and the result was that the model showed higher accuracy than other algorithms. To find the valuable factors of bike rental prediction, the article discussed the influence of holiday, weather and temperature and decided to build the dataset with these significant factors. The article set "di" as 1, 2 and 3 representing weekday, the first half of holiday and the second half of holiday, "wi" as 1, 2 and 3 as different weather and "ti" as different ranges of temperature. To classify different situation of bike rental, the authors used cluster analysis to analyze the modes of bike rental record in training dataset. Through K-means cluster analysis, the bike rental behaviors were divided into 4 modes. Then, the researchers built the classifier using training dataset through Bayes classification method. Then, the article designed an APSO-BP neural network model to forecast the bike rental demand. APSO-BP used single hidden layer neural network model. APSO algorithm generated several particles which are feasible answers randomly at first. The algorithm used the squared error to evaluate fitness. The researchers trained separated ASPO-BP neural network for each mode. The experiment of this research was based on the bike rental record and the weather data in Hangzhou from March 18 to June 15 in 2016. 5 minutes were chosen as the time range. The researchers categorize the training dataset into 4 modes through K-means cluster analysis and then trained ASPO-BP neural network for each mode. The result showed that the accuracy of the demand predicted by this method was influenced by the accuracy of classification and the accuracy of this method in predicting rental demand was higher than other common methods. From this research, I learned that the factors should be transferred into reasonable values firstly and the method of transferring the data are diverse. Besides, the simple neural network models sometimes cannot solve the problem accurately. The combination of different algorithm are needed. Moreover, there are many optimized methods thus researchers should choose the most suitable optimized method to improve the accuracy. These tips can be useful for our project.

This manuscript is a template (aka "rootstock") for Manubot, a tool for writing scholarly manuscripts. Use this template as a starting point for your manuscript.

The rest of this document is a full list of formatting elements/features supported by Manubot. Compare the input (`.md` files in the `/content` directory) to the output you see below.

Data Analysis

Bold text

Semi-bold text

Centered text

Right-aligned text

Italic text

Combined *italics and **bold***

~~Strikethrough~~

1. Ordered list item

2. Ordered list item
    a. Sub-item
    b. Sub-item
        i. Sub-sub-item
3. Ordered list item
    a. Sub-item

- List item
- List item
- List item

subscript: H_2O is a liquid

superscript: 2^{10} is 1024.

unicode superscripts^{0123456789}

unicode subscripts$_{0123456789}$

A long paragraph of text. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Putting each sentence on its own line has numerous benefits with regard to editing and version control.

Line break without starting a new paragraph by putting
two spaces at end of line.

Heading 3

Heading 4

Heading 5

Heading 6

A heading centered on its own printed page

Horizontal rule:

`Heading 3`'s and `Heading 4`'s are recommended for sub-sections.

Links

Bare URL link: https://manubot.org

Long link with lots of words and stuff and junk and bleep and blah and stuff and other stuff and more stuff yeah

Link with text

Link with hover text

Link by reference

Citations

Citation by DOI [1].

Citation by PubMed Central ID [2].

Citation by PubMed ID [3].

Citation by Wikidata ID [4].

Citation by ISBN [5].

Citation by URL [6].

Citation by alias [7].

Multiple citations can be put inside the same set of brackets [1,5,7]. Manubot plugins provide easier, more convenient visualization of and navigation between citations [2,3,7,8].

Citation tags (i.e. aliases) can be defined in their own paragraphs using Markdown's reference link syntax:

Referencing figures, tables, equations

Figure 17

Figure 14

Figure 15

Figure 16

Table 7

Equation 7

Equation 8

Quotes and code

> Quoted text

> Quoted block of text
>
> Two roads diverged in a wood, and I—
> I took the one less traveled by,
> And that has made all the difference.

Code `in the middle` of normal text, aka `inline code`.

Code block with Python syntax highlighting:

```python
from manubot.cite.doi import expand_short_doi

def test_expand_short_doi():
    doi = expand_short_doi("10/c3bp")
    # a string too long to fit within page:
    assert doi == "10.25313/2524-2695-2018-3-vliyanie-enhansera-copia-i-
        insulyatora-gypsy-na-sintez-ernk-modifikatsii-hromatina-i-
        svyazyvanie-insulyatornyh-belkov-vtransfetsirovannyh-geneticheskih-
        konstruktsiyah"
```

Code block with no syntax highlighting:

```
Exporting HTML manuscript
Exporting DOCX manuscript
Exporting PDF manuscript
```

Figures

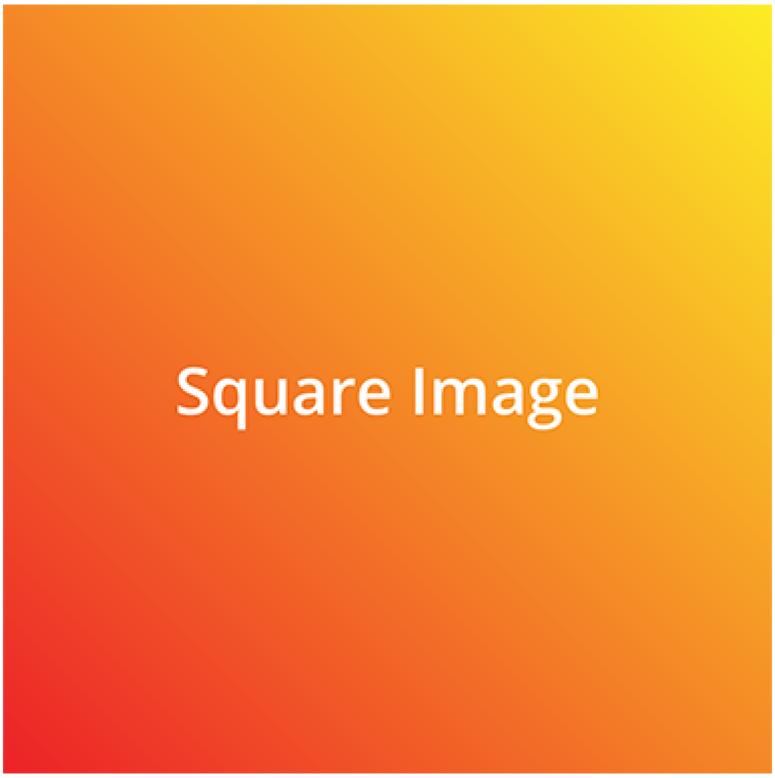

Figure 1: A square image at actual size and with a bottom caption. Loaded from the latest version of image on GitHub.

Figure 2: An image too wide to fit within page at full size. Loaded from a specific (hashed) version of the image on GitHub.

Figure 3: A tall image with a specified height. Loaded from a specific (hashed) version of the image on GitHub.

Figure 4: A vector `.svg` image loaded from GitHub. The parameter `sanitize=true` is necessary to properly load SVGs hosted via GitHub URLs. White background specified to serve as a backdrop for transparent sections of the image.

Tables

Table 1: A table with a top caption and specified relative column widths.

Bowling Scores	**Jane**	**John**	**Alice**	**Bob**
Game 1	150	187	210	105
Game 2	98	202	197	102
Game 3	123	180	238	134

Table 2: A table too wide to fit within page.

	Digits 1-33	**Digits 34-66**	**Digits 67-99**	**Ref.**
pi	3.14159265358979323 846264338327950	28841971693993751 0582097494459230	78164062862089986 2803482534211706	`piday.org`
e	2.71828182845904523 536028747135266	24977572470936999 5957496696762772	40766303535475945 7138217852516642	`nasa.gov`

Table 3: A table with merged cells using the `attributes` plugin.

	Colors	
Size	**Text Color**	**Background Color**
big	blue	orange
small	black	white

Equations

A LaTeX equation:

$$\int_0^\infty e^{-x^2} dx = \frac{\sqrt{\pi}}{2} \tag{1}$$

An equation too long to fit within page:

$$\begin{split} x = a + b + c + d + e + f + g + h + i + j + k + l + m + n + o + p + q + r + s + t \\ + u + v + w + x + y + z + 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 \end{split} \tag{2}$$

Special

⚠ WARNING *The following features are only supported and intended for `.html` and `.pdf` exports. Journals are not likely to support them, and they may not display correctly when converted to other formats such as `.docx`.*

LINK STYLED AS A BUTTON

Adding arbitrary HTML attributes to an element using Pandoc's attribute syntax:

> Manubot Manubot Manubot Manubot Manubot. Manubot Manubot Manubot Manubot. Manubot Manubot Manubot. Manubot Manubot. Manubot.

Adding arbitrary HTML attributes to an element with the Manubot `attributes` plugin (more flexible than Pandoc's method in terms of which elements you can add attributes to):

> Manubot Manubot Manubot Manubot Manubot. Manubot Manubot Manubot Manubot. Manubot Manubot Manubot. Manubot Manubot. Manubot.

Available background colors for text, images, code, banners, etc:

`white` `lightgrey` `grey` `darkgrey` `black` `lightred` `lightyellow` `lightgreen` `lightblue` `lightpurple` `red` `orange` `yellow` `green` `blue` `purple`

Using the Font Awesome icon set:

✔ ? ★ 🔔 ⊗ ⋯

Light Grey Banner
useful for *general information* - manubot.org

Blue Banner
useful for *important information* - manubot.org

Light Red Banner
useful for *warnings* - manubot.org

The Models to Predict the Amount of Bike Sharing

Regular Neural Network

Based on the exploratory data analysis, regular neural network is used to figure out the project.

Neural network is a model that optimize the parameters through learning process to recognize hidden relationships between different data.

Because of the low correlation between given features, regular neural network probably is the most appropriate model to solve the project. The architecture of regular neural network is shown in Figure 5.

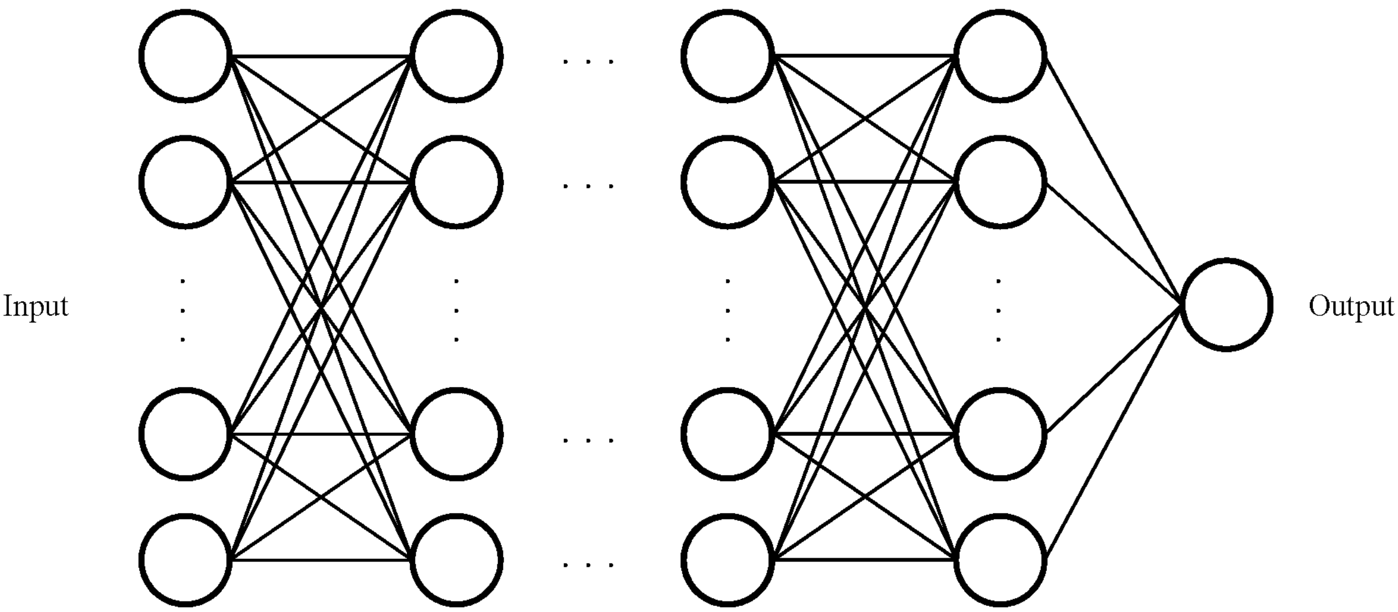

Figure 5: The Architecture of Regular Neural Network

There are 4 steps to build and train neural network, including:

- Selecting features;
- Data preprocessing;
- Designing layers and parameters;
- Determining training methods.

We used different features, epochs, hidden layers, units and learning rates in this project. The evaluation of the models' performances are based on the root mean squared error(RMSE) between the test data and predictions. The architectures of our models are shown in Figure 6, 7 and 8.

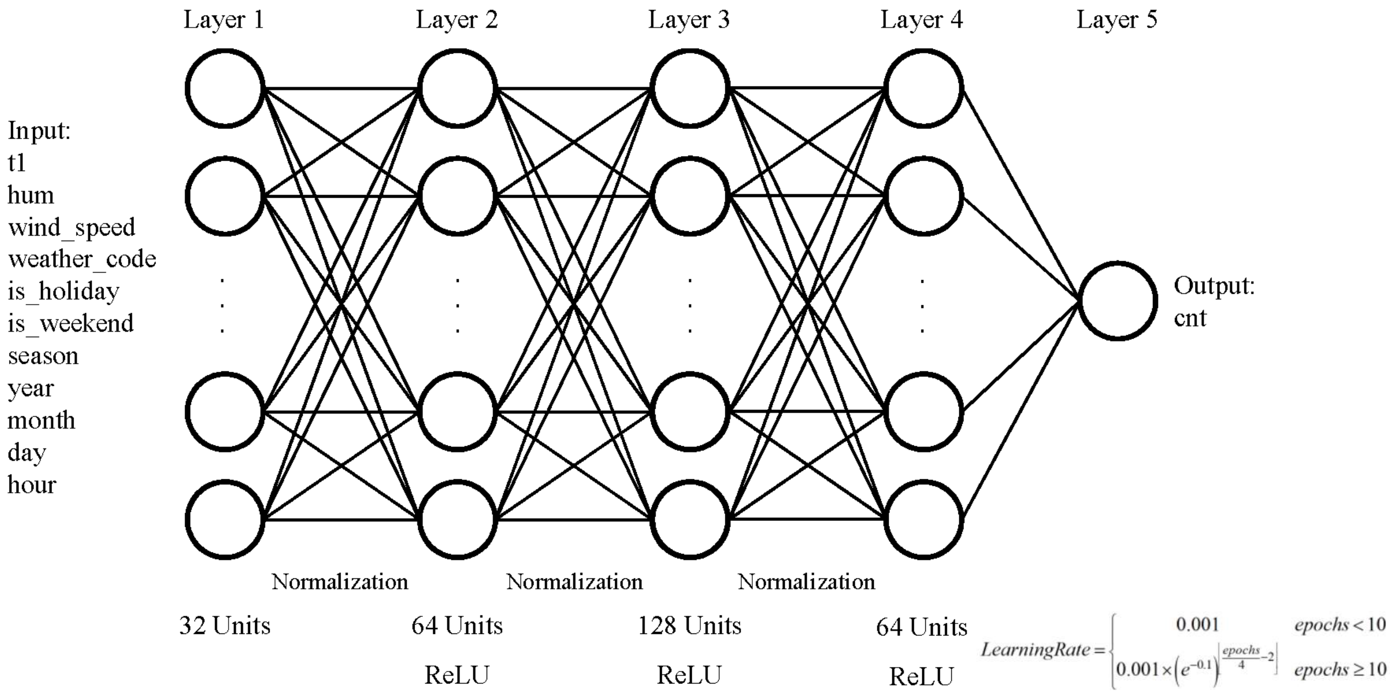

Figure 6: The Architecture of Jingzi's Model

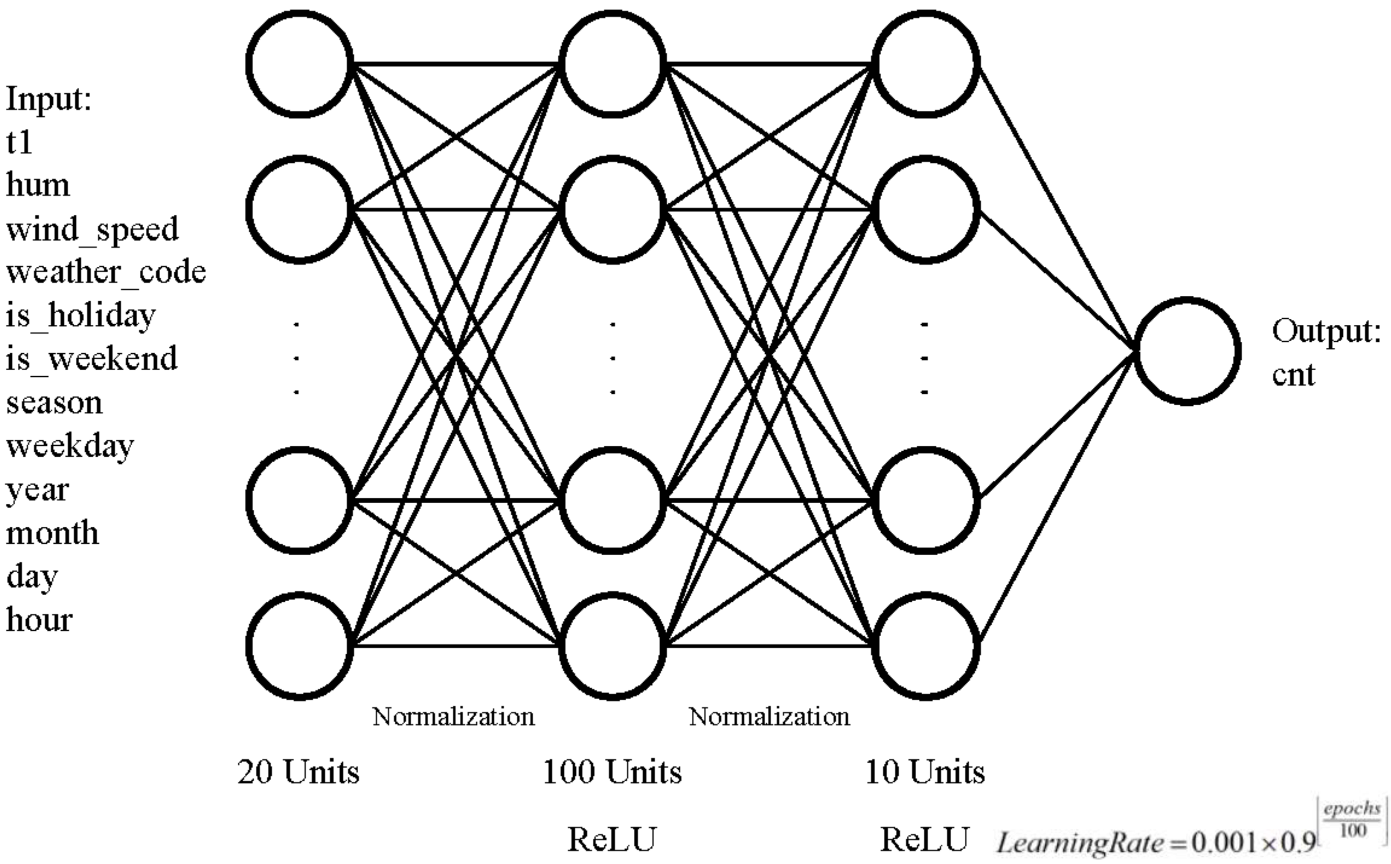

Figure 7: The Architecture of Anye's Model

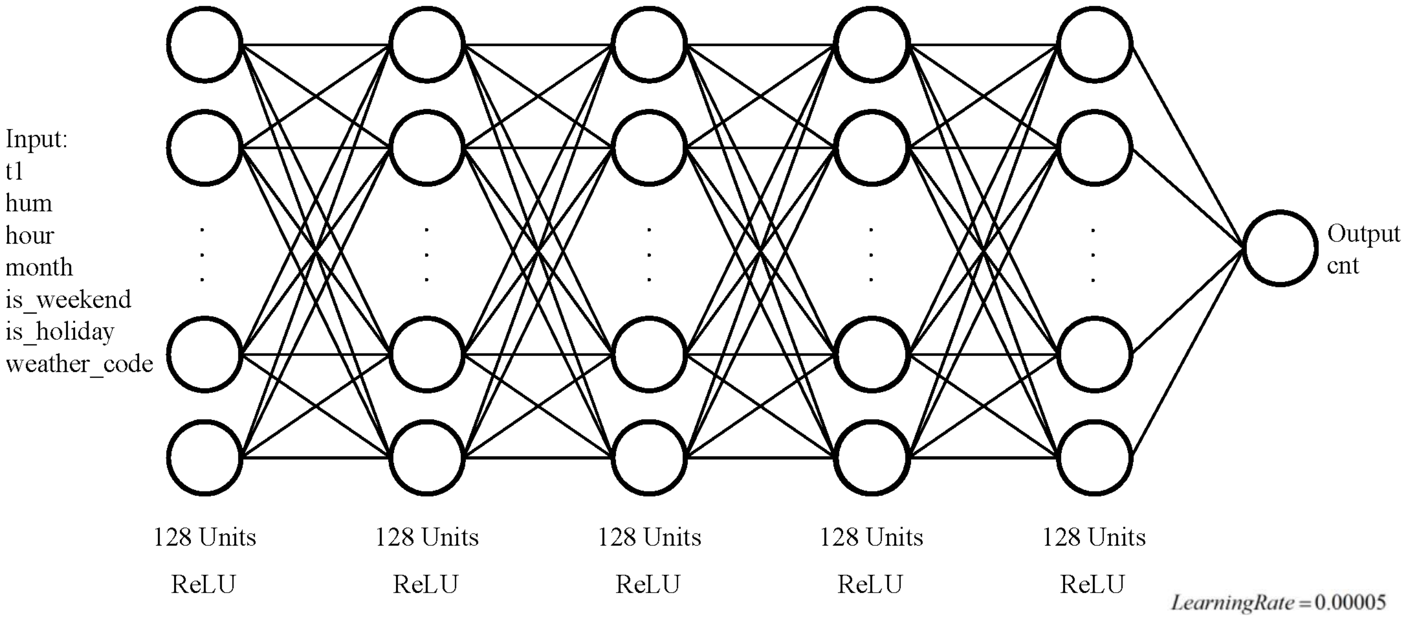

Figure 8: The Architecture of Dana’s Model

Sensitivity Analysis

To evaluate and optimize our models, we analyzed the sensitivity of different parameters.

Units of Layers, Layers, Normalization and Learning Rates are analyzed. To avoid the influence of randomization, the evaluation of the performances is according to the average RMSE of 3 separate training with same initial parameters.

Bold text

Semi-bold text

Centered text

Right-aligned text

Italic text

Combined *italics and **bold***

~~Strikethrough~~

1. Ordered list item
2. Ordered list item
    a. Sub-item
    b. Sub-item
        i. Sub-sub-item
3. Ordered list item
    a. Sub-item

- List item
- List item
- List item

subscript: H_2O is a liquid

superscript: 2^{10} is 1024.

unicode superscripts^{0123456789}

unicode subscripts$_{0123456789}$

A long paragraph of text. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Putting each sentence on its own line has numerous benefits with regard to editing and version control.

Line break without starting a new paragraph by putting
two spaces at end of line.

Heading 3

Heading 4

Heading 5

Heading 6

A heading centered on its own printed page

Horizontal rule:

`Heading 3`'s and `Heading 4`'s are recommended for sub-sections.

Links

Bare URL link: https://manubot.org

Long link with lots of words and stuff and junk and bleep and blah and stuff and other stuff and more stuff yeah

Link with text

Link with hover text

Link by reference

Citations

Citation by DOI [1].

Citation by PubMed Central ID [2].

Citation by PubMed ID [3].

Citation by Wikidata ID [4].

Citation by ISBN [5].

Citation by URL [6].

Citation by alias [7].

Multiple citations can be put inside the same set of brackets [1,5,7]. Manubot plugins provide easier, more convenient visualization of and navigation between citations [2,3,7,8].

Citation tags (i.e. aliases) can be defined in their own paragraphs using Markdown's reference link syntax:

Referencing figures, tables, equations

Figure 17

Figure 14

Figure 15

Figure 16

Table 7

Equation 7

Equation 8

Quotes and code

> Quoted text

> Quoted block of text
>
> Two roads diverged in a wood, and I—
> I took the one less traveled by,
> And that has made all the difference.

Code `in the middle` of normal text, aka `inline code`.

Code block with Python syntax highlighting:

```python
from manubot.cite.doi import expand_short_doi

def test_expand_short_doi():
    doi = expand_short_doi("10/c3bp")
    # a string too long to fit within page:
    assert doi == "10.25313/2524-2695-2018-3-vliyanie-enhansera-copia-i-
        insulyatora-gypsy-na-sintez-ernk-modifikatsii-hromatina-i-
        svyazyvanie-insulyatornyh-belkov-vtransfetsirovannyh-geneticheskih-
        konstruktsiyah"
```

Code block with no syntax highlighting:

```
Exporting HTML manuscript
Exporting DOCX manuscript
Exporting PDF manuscript
```

Figures

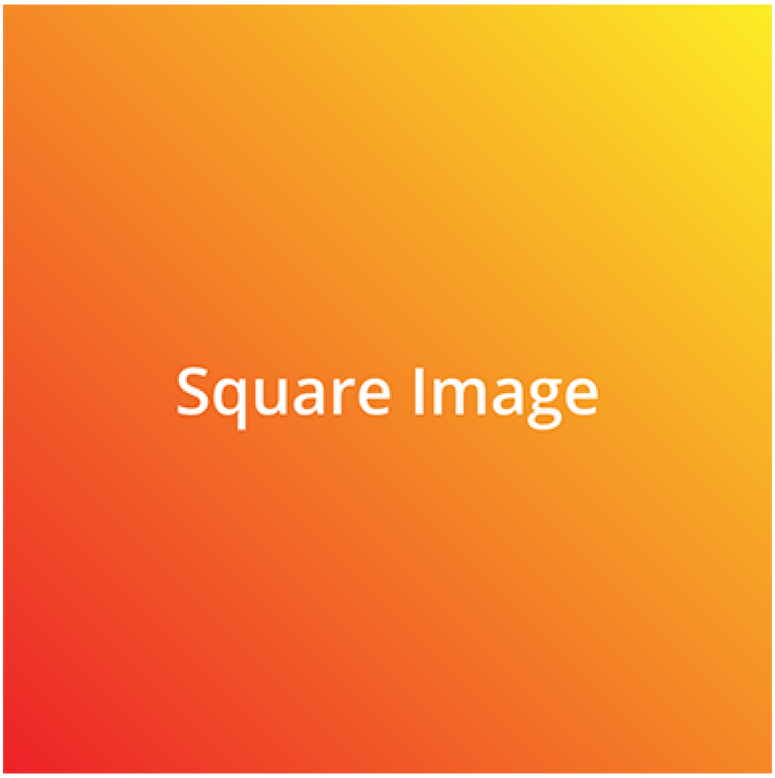

Figure 9: A square image at actual size and with a bottom caption. Loaded from the latest version of image on GitHub.

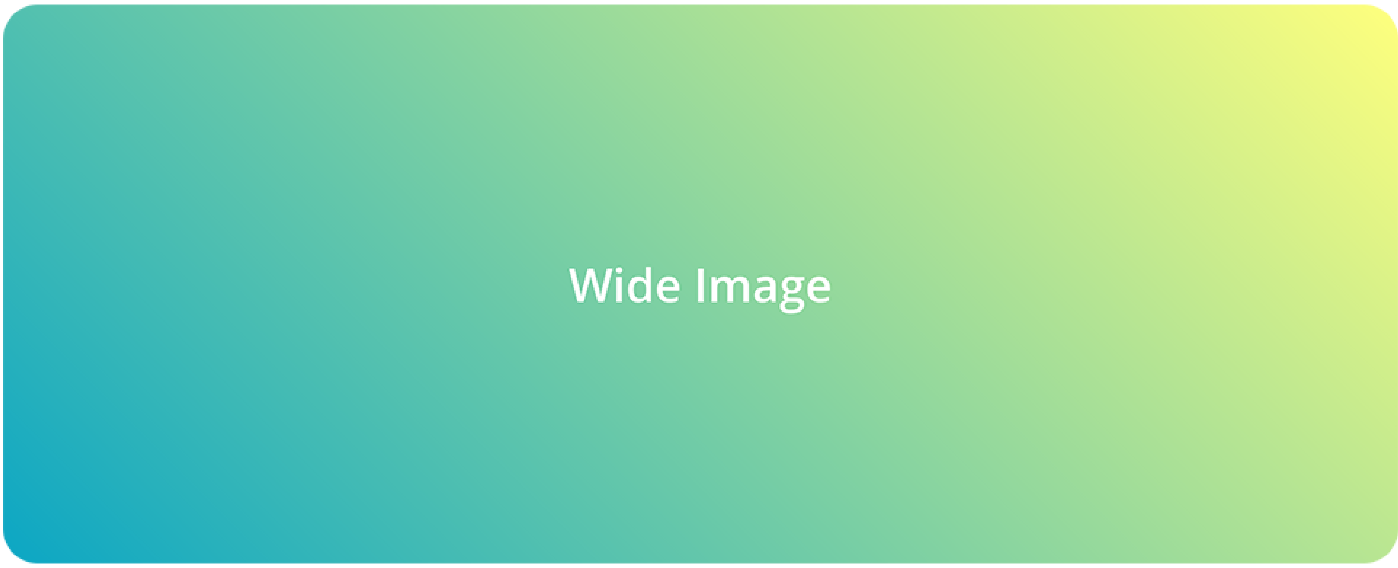

Figure 10: An image too wide to fit within page at full size. Loaded from a specific (hashed) version of the image on GitHub.

Figure 11: A tall image with a specified height. Loaded from a specific (hashed) version of the image on GitHub.

Figure 12: A vector `.svg` image loaded from GitHub. The parameter `sanitize=true` is necessary to properly load SVGs hosted via GitHub URLs. White background specified to serve as a backdrop for transparent sections of the image.

Tables

Table 4: A table with a top caption and specified relative column widths.

Bowling Scores	**Jane**	**John**	**Alice**	**Bob**
Game 1	150	187	210	105
Game 2	98	202	197	102
Game 3	123	180	238	134

Table 5: A table too wide to fit within page.

	Digits 1-33	**Digits 34-66**	**Digits 67-99**	**Ref.**
pi	3.14159265358979323 846264338327950	28841971693993751 0582097494459230	78164062862089986 2803482534211706	`piday.org`
e	2.71828182845904523 536028747135266	24977572470936999 5957496696762772	40766303535475945 7138217852516642	`nasa.gov`

Table 6: A table with merged cells using the `attributes` plugin.

	Colors	
Size	**Text Color**	**Background Color**
big	blue	orange
small	black	white

Equations

A LaTeX equation:

$$\int_0^\infty e^{-x^2} dx = \frac{\sqrt{\pi}}{2} \tag{3}$$

An equation too long to fit within page:

$$\begin{aligned} x = a + b + c + d + e + f + g + h + i + j + k + l + m + n + o + p + q + r + s + t \\ + u + v + w + x + y + z + 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 \end{aligned} \tag{4}$$

Special

⚠ WARNING *The following features are only supported and intended for `.html` and `.pdf` exports. Journals are not likely to support them, and they may not display correctly when converted to other formats such as `.docx`.*

LINK STYLED AS A BUTTON

Adding arbitrary HTML attributes to an element using Pandoc's attribute syntax:

> Manubot Manubot Manubot Manubot Manubot. Manubot Manubot Manubot Manubot. Manubot Manubot Manubot. Manubot Manubot. Manubot.

Adding arbitrary HTML attributes to an element with the Manubot `attributes` plugin (more flexible than Pandoc's method in terms of which elements you can add attributes to):

> Manubot Manubot Manubot Manubot Manubot. Manubot Manubot Manubot Manubot. Manubot Manubot Manubot. Manubot Manubot. Manubot.

Available background colors for text, images, code, banners, etc:

`white` `lightgrey` `grey` `darkgrey` `black` `lightred` `lightyellow` `lightgreen` `lightblue` `lightpurple` `red` `orange` `yellow` `green` `blue` `purple`

Using the [Font Awesome](https://fontawesome.com/) icon set:

✔ ? ★ 🔔 ⊗ ⋯

Light Grey Banner
useful for *general information* - manubot.org

Blue Banner
useful for *important information* - manubot.org

Light Red Banner
useful for *warnings* - manubot.org

This manuscript is a template (aka "rootstock") for Manubot, a tool for writing scholarly manuscripts. Use this template as a starting point for your manuscript.

The rest of this document is a full list of formatting elements/features supported by Manubot. Compare the input (`.md` files in the `/content` directory) to the output you see below.

Experiments of the Prediction Models

Bold text

Semi-bold text

Centered text

Right-aligned text

Italic text

Combined *italics and **bold***

~~Strikethrough~~

1. Ordered list item
2. Ordered list item
   a. Sub-item
   b. Sub-item
      i. Sub-sub-item
3. Ordered list item
   a. Sub-item

- List item
- List item
- List item

subscript: H_2O is a liquid

superscript: 2^{10} is 1024.

unicode superscripts^{0123456789}

unicode subscripts$_{0123456789}$

A long paragraph of text. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Putting each sentence on its own line has numerous benefits with regard to editing and version control.

Line break without starting a new paragraph by putting
two spaces at end of line.

Heading 3

Heading 4

Heading 5

Heading 6

A heading centered on its own printed page

Horizontal rule:

`Heading 3`'s and `Heading 4`'s are recommended for sub-sections.

Links

Bare URL link: https://manubot.org

Long link with lots of words and stuff and junk and bleep and blah and stuff and other stuff and more stuff yeah

Link with text

Link with hover text

Link by reference

Citations

Citation by DOI [1].

Citation by PubMed Central ID [2].

Citation by PubMed ID [3].

Citation by Wikidata ID [4].

Citation by ISBN [5].

Citation by URL [6].

Citation by alias [7].

Multiple citations can be put inside the same set of brackets [1,5,7]. Manubot plugins provide easier, more convenient visualization of and navigation between citations [2,3,7,8].

Citation tags (i.e. aliases) can be defined in their own paragraphs using Markdown's reference link syntax:

Referencing figures, tables, equations

Figure 17

Figure 14

Figure 15

Figure 16

Table 7

Equation 7

Equation 8

Quotes and code

> Quoted text

> Quoted block of text
>
> Two roads diverged in a wood, and I—
> I took the one less traveled by,
> And that has made all the difference.

Code `in the middle` of normal text, aka `inline code`.

Code block with Python syntax highlighting:

```python
from manubot.cite.doi import expand_short_doi

def test_expand_short_doi():
    doi = expand_short_doi("10/c3bp")
    # a string too long to fit within page:
    assert doi == "10.25313/2524-2695-2018-3-vliyanie-enhansera-copia-i-
        insulyatora-gypsy-na-sintez-ernk-modifikatsii-hromatina-i-
        svyazyvanie-insulyatornyh-belkov-vtransfetsirovannyh-geneticheskih-
        konstruktsiyah"
```

Code block with no syntax highlighting:

```
Exporting HTML manuscript
Exporting DOCX manuscript
Exporting PDF manuscript
```

Figures

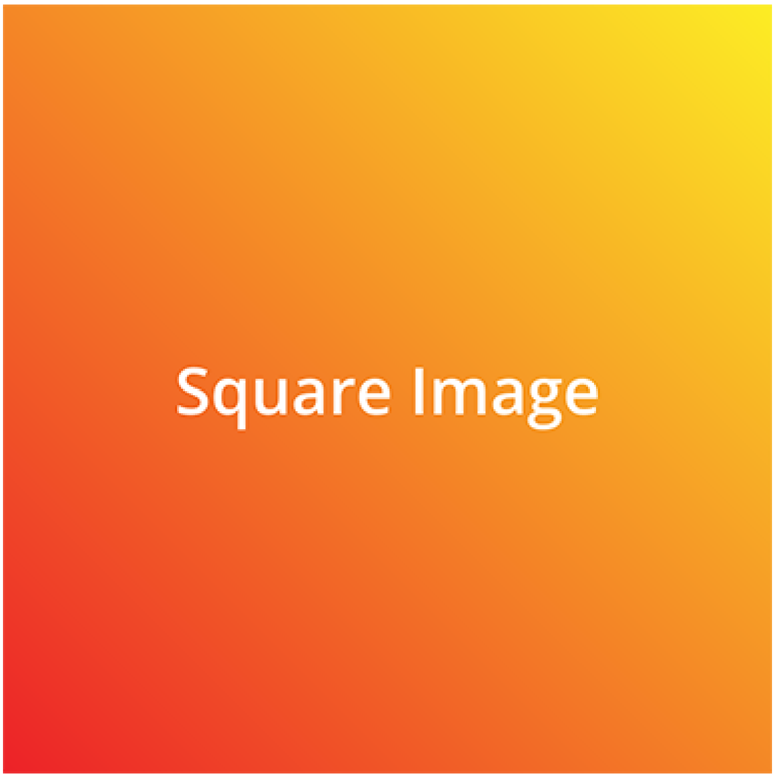

Figure 13: A square image at actual size and with a bottom caption. Loaded from the latest version of image on GitHub.

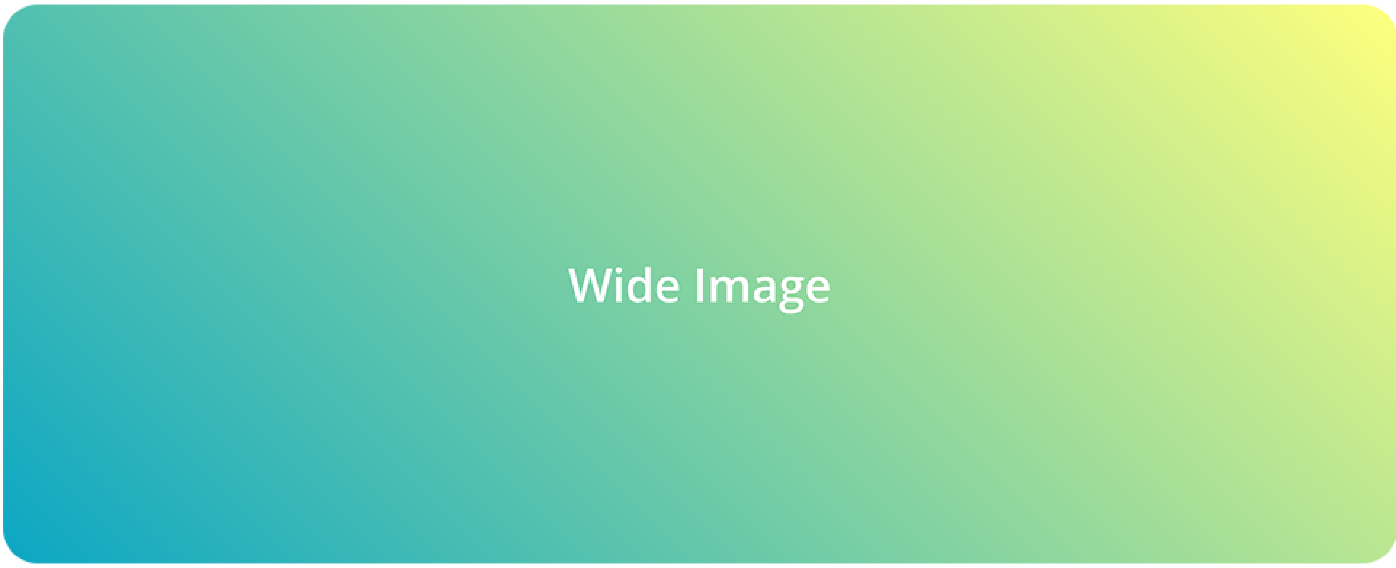

Figure 14: An image too wide to fit within page at full size. Loaded from a specific (hashed) version of the image on GitHub.

Figure 15: A tall image with a specified height. Loaded from a specific (hashed) version of the image on GitHub.

Figure 16: A vector `.svg` image loaded from GitHub. The parameter `sanitize=true` is necessary to properly load SVGs hosted via GitHub URLs. White background specified to serve as a backdrop for transparent sections of the image.

Tables

Table 7: A table with a top caption and specified relative column widths.

Bowling Scores	**Jane**	**John**	**Alice**	**Bob**
Game 1	150	187	210	105
Game 2	98	202	197	102
Game 3	123	180	238	134

Table 8: A table too wide to fit within page.

	Digits 1-33	**Digits 34-66**	**Digits 67-99**	**Ref.**
pi	3.141592653589793238462643383279 50	288419716939937510582097494459230	781640628620899862803482534211706	`piday.org`
e	2.718281828459045235360287471352 66	249775724709369995957496696762772	407663035354759457138217852516642	`nasa.gov`

Table 9: A table with merged cells using the `attributes` plugin.

	Colors	
Size	**Text Color**	**Background Color**
big	blue	orange
small	black	white

Equations

A LaTeX equation:

$$\int_0^\infty e^{-x^2} dx = \frac{\sqrt{\pi}}{2} \tag{5}$$

An equation too long to fit within page:

$$\begin{split} x = a + b + c + d + e + f + g + h + i + j + k + l + m + n + o + p + q + r + s + t \\ + u + v + w + x + y + z + 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 \end{split} \tag{6}$$

Special

⚠ WARNING *The following features are only supported and intended for `.html` and `.pdf` exports. Journals are not likely to support them, and they may not display correctly when converted to other formats such as `.docx`.*

LINK STYLED AS A BUTTON

Adding arbitrary HTML attributes to an element using Pandoc's attribute syntax:

> Manubot Manubot Manubot Manubot Manubot. Manubot Manubot Manubot Manubot. Manubot Manubot Manubot. Manubot Manubot. Manubot.

Adding arbitrary HTML attributes to an element with the Manubot `attributes` plugin (more flexible than Pandoc's method in terms of which elements you can add attributes to):

> Manubot Manubot Manubot Manubot Manubot. Manubot Manubot Manubot Manubot. Manubot Manubot Manubot. Manubot Manubot. Manubot.

Available background colors for text, images, code, banners, etc:

`white` `lightgrey` `grey` `darkgrey` `black` `lightred` `lightyellow` `lightgreen` `lightblue` `lightpurple` `red` `orange` `yellow` `green` `blue` `purple`

Using the [Font Awesome](https://fontawesome.com) icon set:

✔ ? ★ 🔔 ⊗ ⋯

Light Grey Banner
useful for *general information* - manubot.org

Blue Banner
useful for *important information* - manubot.org

Light Red Banner
useful for *warnings* - manubot.org

Discussion

Plot of the actual and predicted bike counts with respect to month. Loaded from the latest version of image on GitHub.

Plot of the actual and predicted bike counts with respect to month. Loaded from the latest version of image on GitHub.

Conclusion

References

[1] Z. Kang, Y. Zuo, Z. Huang, F. Zhou and P. Chen, "Research on the Forecast of Shared Bicycle Rental Demand Based on Spark Machine Learning Framework," 2017 16th International Symposium on Distributed Computing and Applications to Business, Engineering and Science (DCABES), Anyang, 2017, pp. 219-222, doi: 10.1109/DCABES.2017.55.

[2] F. Lin, S. Wang, J. Jiang, W. Fan and Y. Sun, "Predicting public bicycle rental number using multi-source data," 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, 2017, pp. 1502-1509, doi: 10.1109/IJCNN.2017.7966030.

[3] J. Huang, X. Wang and H. Sun, "Central Station Based Demand Prediction in a Bike Sharing System," 2019 20th IEEE International Conference on Mobile Data Management (MDM), Hong Kong, Hong Kong, 2019, pp. 346-348, doi: 10.1109/MDM.2019.00-38.

[4] https://www.kaggle.com/hmavrodiev/london-bike-sharing-dataset

[5] Ashqar, H. I., Elhenawy, M., Rakha, H. A., Road, V. P., & Qld, K. G. "Case Studies on Transport Policy Modeling bike counts in a bike-sharing system considering the e ff ect of weather conditions," 2019. Case Studies on Transport Policy, 7(2), 261–268. https://doi.org/10.1016/j.cstp.2019.02.011

[6] F. Lin, S. Wang, J. Jiang, W. Fan and Y. Sun, "Predicting public bicycle rental number using multi-source data," 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, 2017, pp. 1502-1509, doi: 10.1109/IJCNN.2017.7966030.

Referencing figures, tables, equations

Figure 17

Figure 14

Figure 15

Figure 16

Table 7

Equation 7

Equation 8

Figures

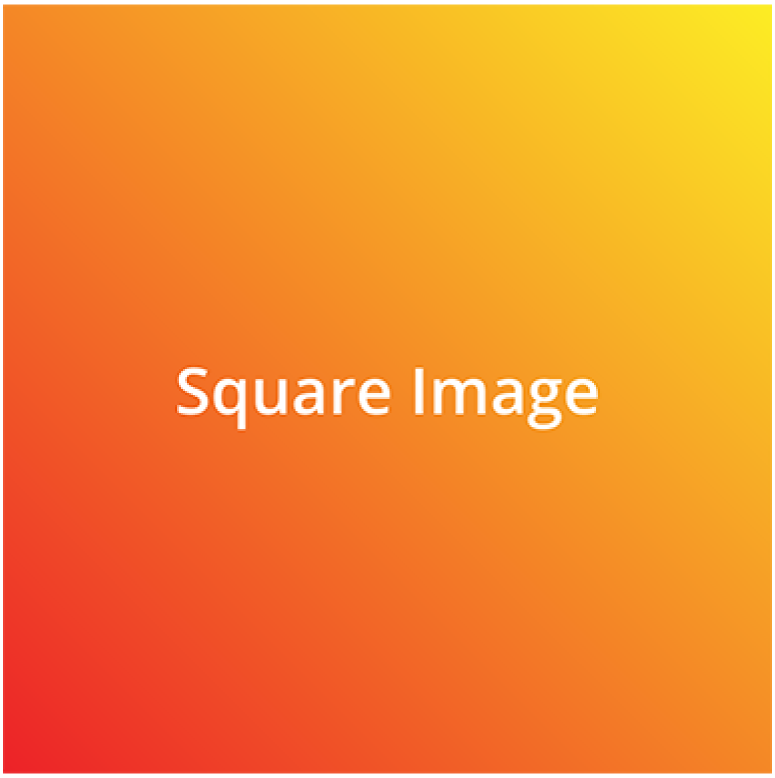

Figure 17: A square image at actual size and with a bottom caption. Loaded from the latest version of image on GitHub.

Equations

A LaTeX equation:

$$\int_0^\infty e^{-x^2} dx = \frac{\sqrt{\pi}}{2} \tag{7}$$

An equation too long to fit within page:

$$x = a + b + c + d + e + f + g + h + i + j + k + l + m + n + o + p + q + r + s + t \\ + u + v + w + x + y + z + 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 \tag{8}$$

References

1. **Sci-Hub provides access to nearly all scholarly literature**
Daniel S Himmelstein, Ariel Rodriguez Romero, Jacob G Levernier, Thomas Anthony Munro, Stephen Reid McLaughlin, Bastian Greshake Tzovaras, Casey S Greene
eLife (2018-03-01) https://doi.org/ckcj
DOI: 10.7554/elife.32822 · PMID: 29424689 · PMCID: PMC5832410

2. **Reproducibility of computational workflows is automated using continuous analysis**
Brett K Beaulieu-Jones, Casey S Greene
Nature biotechnology (2017-04) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6103790/
DOI: 10.1038/nbt.3780 · PMID: 28288103 · PMCID: PMC6103790

3. **Bitcoin for the biological literature.**
Douglas Heaven
Nature (2019-02) https://www.ncbi.nlm.nih.gov/pubmed/30718888
DOI: 10.1038/d41586-019-00447-9 · PMID: 30718888

4. **Plan S: Accelerating the transition to full and immediate Open Access to scientific publications**
cOAlition S
(2018-09-04) https://www.wikidata.org/wiki/Q56458321

5. **Open access**
Peter Suber
MIT Press (2012)
ISBN: 9780262517638

6. **Open collaborative writing with Manubot**
Daniel S. Himmelstein, Vincent Rubinetti, David R. Slochower, Dongbo Hu, Venkat S. Malladi, Casey S. Greene, Anthony Gitter
Manubot (2020-05-25) https://greenelab.github.io/meta-review/

7. **Opportunities and obstacles for deep learning in biology and medicine**
Travers Ching, Daniel S. Himmelstein, Brett K. Beaulieu-Jones, Alexandr A. Kalinin, Brian T. Do, Gregory P. Way, Enrico Ferrero, Paul-Michael Agapow, Michael Zietz, Michael M. Hoffman, … Casey S. Greene
Journal of The Royal Society Interface (2018-04-04) https://doi.org/gddkhn
DOI: 10.1098/rsif.2017.0387 · PMID: 29618526 · PMCID: PMC5938574

8. **Open collaborative writing with Manubot**
Daniel S. Himmelstein, Vincent Rubinetti, David R. Slochower, Dongbo Hu, Venkat S. Malladi, Casey S. Greene, Anthony Gitter
PLOS Computational Biology (2019-06-24) https://doi.org/c7np
DOI: 10.1371/journal.pcbi.1007128 · PMID: 31233491 · PMCID: PMC6611653