No Code ML

Vedant Barbhaya1, Viraj Patel2, Vishal Kundar3, Vidya U4

*Department of Information Science & Engineering,*

*CMR Institute of Technology, Bengaluru.*

1veda17is@cmrit.ac.in,2vipa17is@cmrit.ac.in,

3vish17is@cmrit.ac.in,4vidya.u@cmrit.ac.in

***Abstract*— Data is the oil of the 21st century. Essentially if you are running a service/business, every interaction with a customer generates some kind of valuable data. With the advancements of intelligent systems, it has become possible to extract very valuable information from any given data. But not every business has the ability or means to develop an in-house team to analyse vast amounts of data generated by them.**

**No Code ML is trying to create a platform where a user can upload a data set and get personalized predictions with just one click. This eliminates the need to clean or process the data or create a machine learning model. No code ML helps businesses to analyse their data using ML without actually needing the knowledge of how to code it. Everything will be handled by our platform and the user will get a personalized dashboard to analyse the outcomes.**

***Keywords***— **Machine Learning (ML),**

1. Introduction

With the success of Machine Learning (ML) in recent years, it has started attracting a lot of attention from the research and business communities.

Machine Learning involves the design and development of pipelines for applications and ML systems.

Building such a pipeline requires a team of human experts: data scientists having statistical and ML knowledge; domain experts with years of experience within a specific domain. Together, these human experts can build a sensible ML pipeline containing data preprocessing, meaningful feature engineering, and fine-tuned models leading to great results.

Every machine learning service, at its core, needs to solve the same fundamental problems: deciding which machine learning algorithm to use on a given dataset, whether and how to preprocess its features, how to generate meaningful features, and how to tune all hyperparameters.

This process is a complex task, performed in an iterative manner with trial and error. Building a good ML pipeline is a laborious process and practitioners often use a suboptimal default ML pipeline.

To solve these issues, a novel idea of automating the entire pipeline of machine learning (ML) has emerged, i.e., automated machine learning (AutoML).

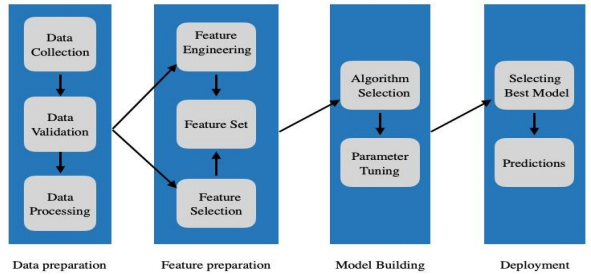
The goal of AutoML is to allow practitioners to build ML applications without much requirement for statistical and ML knowledge.

With the advancements in Cloud and BigData technologies, AutoML has started gaining a lot of attention. A complete AutoML system can dynamically combine various techniques to form an easy-to-use end-to-end ML pipeline system.

In the context of supervised learning, AutoML aims at removing user interaction altogether from all stages of the design and development of supervised learning systems.

As data is being generated at an incredible rate in practically any context and scenario, the number of practitioners available to analyze such data are limited.

AutoML pipeline consists of several processes: data preparation, feature engineering, model generation, and model evaluation. Given below is an illustration of a typical ML pipeline.



1. Related Work

According to the no free lunch theorem (Wolpert and Macready, 1997) it is impossible that a single optimization algorithm is universally superior to any other optimization algorithm. This implies that no universally superior ML pipeline for all ML tasks exists. Consequently, it is not possible to transfer a well performing ML pipeline to a new domain and still yield outstanding results. Instead, a new ML pipeline has to be constructed for each new data set to obtain optimal results. However, manually building a specialized ML pipeline for each and every problem is very time consuming and therefore expensive. As a consequence, practitioners often use a suboptimal default ML pipeline.

AutoML aims to improve the current way of building ML applications by automation. ML experts can profit from AutoML by automating tedious tasks like Hyperparameter optimization leading to a higher efficiency. Domain experts can be enabled to build ML pipelines on their own without having to rely on a data scientist.

It is important to note that AutoML is no new trend. Starting from the 1990s commercial solutions offered automatic Hyperparameter optimization for selected classification algorithms via grid search (Dinsmore, 2016). In 2004, the first efficient strategies for Hyperparameter optimization were proposed. For limited settings, e.g., tuning C and γ of a SVM (Chen et al., 2004), it was proven that guided search strategies yield better results than grid search in less time. Also in 2004, the first approaches for automatic feature selection were published (Samanta, 2004). Full model selection (Escalante et al., 2009) was the first attempt to automatically build a complete ML pipeline by simultaneously selecting a preprocessing, feature selection and classification algorithm while tuning the hyperparameters of each method. Testing this approach on various data sets, the potential of this domain-agnostic method was proven (Guyon et al., 2008). Starting from 2011, many different methods of applying Bayesian optimization for hyperparameter tuning (Bergstra et al., 2011; Snoek et al., 2012) and model selection (Thornton et al., 2013) have been proposed. In 2015, the first method for automatic feature engineering without domain knowledge was proposed (Kanter and Veeramachaneni, 2015). Building arbitrary shaped pipelines has been possible since 2016 (Olson and Moore, 2016). In 2017 and 2018 the topic AutoML received a lot of attention in media (Google, 2019) with the release of commercial AutoML solutions from various global players (Golovin et al., 2017; Clouder, 2018; Baidu, 2018). Simultaneously, research in the area of AutoML gained significant traction leading to many performance improvements. Recent methods are able to reduce the runtime of AutoML procedures from several hours to mere minutes (Hutter et al., 2018). To further optimize AutoML for deep learning, a new stack called LEAF stack was developed which can optimize parameters, components, and topology of the architecture simultaneously to fit the requirements faster than the current state-of-the-art hand designed architectures (Jason Liang et al., 2019)

1. Page Style

All paragraphs must be indented. All paragraphs must be justified, i.e. both left-justified and right-justified.

1. *Text Font of Entire Document*

The entire document should be in Times New Roman or Times font. Type 3 fonts must not be used. Other font types may be used if needed for special purposes.

Recommended font sizes are shown in Table 1.

1. *Title and Author Details*

Title must be in 24 pt Regular font. Author name must be in 11 pt Regular font. Author affiliation must be in 10 pt Italic. Email address must be in 9 pt Courier Regular font.

TABLE I  
Font Sizes for Papers

|  |  |  |  |
| --- | --- | --- | --- |
| **Font Size** | **Appearance (in Time New Roman or Times)** | | |
| **Regular** | **Bold** | **Italic** |
| 8 | table caption (in Small Caps),  figure caption,  reference item |  | reference item (partial) |
| 9 | author email address (in Courier),  cell in a table | abstract body | abstract heading (also in Bold) |
| 10 | level-1 heading (in Small Caps),  paragraph |  | level-2 heading,  level-3 heading,  author affiliation |
| 11 | author name |  |  |
| 24 | title |  |  |

All title and author details must be in single-column format and must be centered.

Every word in a title must be capitalized except for short minor words such as “a”, “an”, “and”, “as”, “at”, “by”, “for”, “from”, “if”, “in”, “into”, “on”, “or”, “of”, “the”, “to”, “with”.

Author details must not show any professional title (e.g. Managing Director), any academic title (e.g. Dr.) or any membership of any professional organization (e.g. Senior Member IEEE).

To avoid confusion, the family name must be written as the last part of each author name (e.g. John A.K. Smith).

Each affiliation must include, at the very least, the name of the company and the name of the country where the author is based (e.g. Causal Productions Pty Ltd, Australia).

Email address is compulsory for the corresponding author.

1. *Section Headings*

No more than 3 levels of headings should be used. All headings must be in 10pt font. Every word in a heading must be capitalized except for short minor words as listed in Section III-B.

1. *Level-1 Heading*: A level-1 heading must be in Small Caps, centered and numbered using uppercase Roman numerals. For example, see heading “III. Page Style” of this document. The two level-1 headings which must not be numbered are “Acknowledgment” and “References”.
2. *Level-2 Heading:* A level-2 heading must be in Italic, left-justified and numbered using an uppercase alphabetic letter followed by a period. For example, see heading “C. Section Headings” above.
3. *Level-3 Heading:* A level-3 heading must be indented, in Italic and numbered with an Arabic numeral followed by a right parenthesis. The level-3 heading must end with a colon. The body of the level-3 section immediately follows the level-3 heading in the same paragraph. For example, this paragraph begins with a level-3 heading.
4. *Figures and Tables*

Figures and tables must be centered in the column. Large figures and tables may span across both columns. Any table or figure that takes up more than 1 column width must be positioned either at the top or at the bottom of the page.

Graphics may be full color. All colors will be retained on the CDROM. Graphics must not use stipple fill patterns because they may not be reproduced properly. Please use only *SOLID FILL* colors which contrast well both on screen and on a black-and-white hardcopy, as shown in Fig. 1.

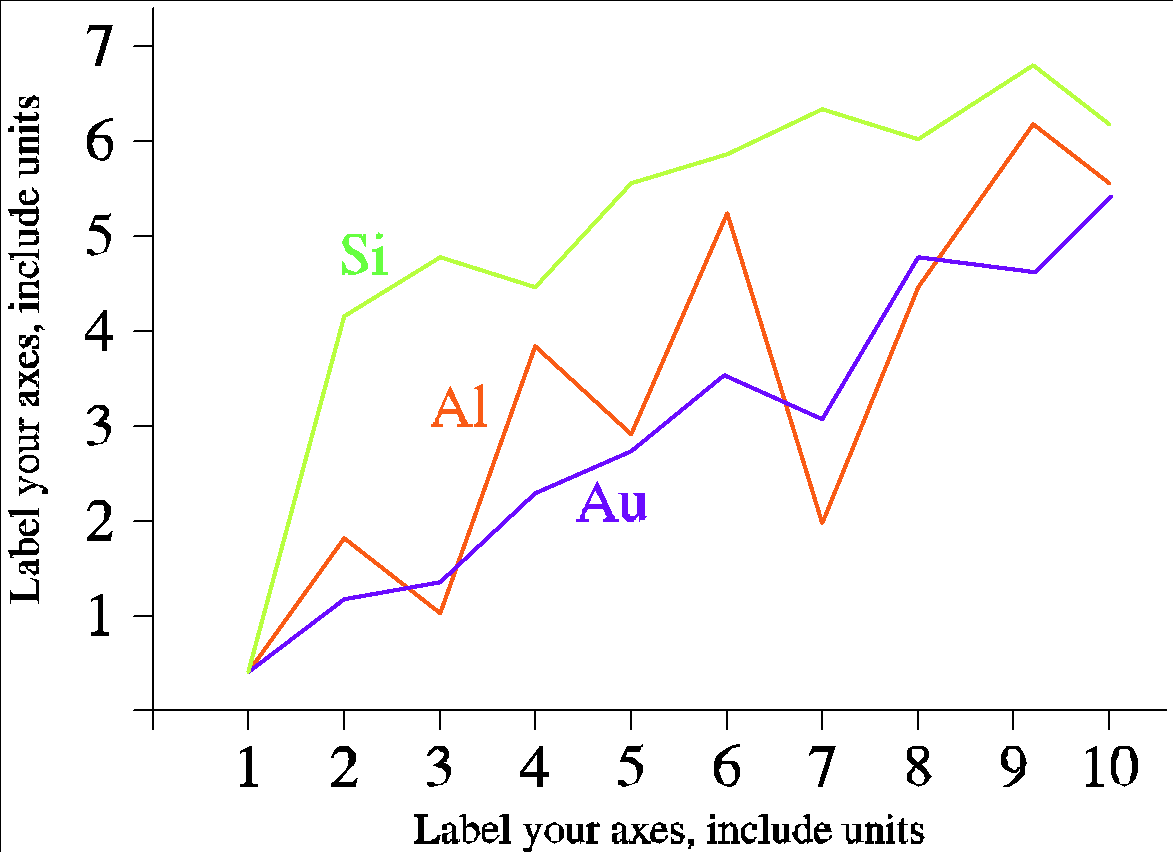


Fig. 1 A sample line graph using colors which contrast well both on screen and on a black-and-white hardcopy

Fig. 2 shows an example of a low-resolution image which would not be acceptable, whereas Fig. 3 shows an example of an image with adequate resolution. Check that the resolution is adequate to reveal the important detail in the figure.

Please check all figures in your paper both on screen and on a black-and-white hardcopy. When you check your paper on a black-and-white hardcopy, please ensure that:

* the colors used in each figure contrast well,
* the image used in each figure is clear,
* all text labels in each figure are legible.

1. *Figure Captions*

Figures must be numbered using Arabic numerals. Figure captions must be in 8 pt Regular font. Captions of a single line (e.g. Fig. 2) must be centered whereas multi-line captions must be justified (e.g. Fig. 1). Captions with figure numbers must be placed after their associated figures, as shown in Fig. 1.



Fig. 2 Example of an unacceptable low-resolution image



Fig. 3 Example of an image with acceptable resolution

1. *Table Captions*

Tables must be numbered using uppercase Roman numerals. Table captions must be centred and in 8 pt Regular font with Small Caps. Every word in a table caption must be capitalized except for short minor words as listed in Section III-B. Captions with table numbers must be placed before their associated tables, as shown in Table 1.

1. *Page Numbers, Headers and Footers*

Page numbers, headers and footers must not be used.

1. *Links and Bookmarks*

All hypertext links and section bookmarks will be removed from papers during the processing of papers for publication. If you need to refer to an Internet email address or URL in your paper, you must type out the address or URL fully in Regular font.

1. *References*

The heading of the References section must not be numbered. All reference items must be in 8 pt font. Please use Regular and Italic styles to distinguish different fields as shown in the References section. Number the reference items consecutively in square brackets (e.g. [1]).

When referring to a reference item, please simply use the reference number, as in [2]. Do not use “Ref. [3]” or “Reference [3]” except at the beginning of a sentence, e.g. “Reference [3] shows …”. Multiple references are each numbered with separate brackets (e.g. [2], [3], [4]–[6]).

Examples of reference items of different categories shown in the References section include:

* example of a book in [1]
* example of a book in a series in [2]
* example of a journal article in [3]
* example of a conference paper in [4]
* example of a patent in [5]
* example of a website in [6]
* example of a web page in [7]
* example of a databook as a manual in [8]
* example of a datasheet in [9]
* example of a master’s thesis in [10]
* example of a technical report in [11]
* example of a standard in [12]

1. Conclusions

The version of this template is V2. Most of the formatting instructions in this document have been compiled by Causal Productions from the IEEE LaTeX style files. Causal Productions offers both A4 templates and US Letter templates for LaTeX and Microsoft Word. The LaTeX templates depend on the official IEEEtran.cls and IEEEtran.bst files, whereas the Microsoft Word templates are self-contained. Causal Productions has used its best efforts to ensure that the templates have the same appearance.

Causal Productions permits the distribution and revision of these templates on the condition that Causal Productions is credited in the revised template as follows: “original version of this template was provided by courtesy of Causal Productions (www.causalproductions.com)”.

Acknowledgment

The heading of the Acknowledgment section and the References section must not be numbered.

Causal Productions wishes to acknowledge Michael Shell and other contributors for developing and maintaining the IEEE LaTeX style files which have been used in the preparation of this template. To see the list of contributors, please refer to the top of file IEEETran.cls in the IEEE LaTeX distribution.

References

1. Automated Machine Learning - a brief review at the end of the early years: Hugo Jair Escalante
2. Efficient and Robust Automated Machine Learning: Matthias Feurer, Aaron Klein Katharina
3. Eggensperger, Jost Tobias Springenberg, Manuel Blum, Frank Hutter
4. Automated Machine Learning: State-of-The-Art and Open Challenges: Radwa Elshawi,
5. Mohamed Maher, Sherif Sakr
6. AutoML: A Survey of the State-of-the-Art: Xin He, Kaiyong Zhao, Xiaowen Chu
7. Benchmark and Survey of Automated Machine Learning Frameworks: Marc-Andr´e Z¨oller, Marco F. Huber