**Machine Learning Model Builder**

Shital Sobale, Harshwardhan More

***Abstract* — *For any machine learning application to be implemented on any mega-scale an automated system is required. In this era of digitization and a technology-driven world, we require software that saves time and does repetitive things. The machine learning model builder discussed in the paper is capable of doing repetitive tasks and generates code in a few clicks. Our architecture is designed on the fundamentals of machine learning and the python-based framework - Django.***

***Keywords – Machine Learning; Data Science; Algorithms; Data Analysis;***

# **Introduction**

In the field of machine learning, students used to do import to export code from scratch. Python and R are programming languages both promote the trending technology known as “Machine learning”. Machine learning is a technology that allows machines to learn from data and do predictive tasks without being explicitly programmed. This web application builds a machine learning model on a dataset that is provided by the user. The necessary code for machine learning model training is written at the backend, the user just has to choose perfect options and click appropriate buttons to build the model and code is generated for service. Users can view uploaded datasets as well as data details along with like mean, median, etc. Also, the web application provides certain features like features relationship, covariance, and feature dependency. Users can see a few relevant visualizations. Based on these few data information, users can easily train machine learning models on best fit recommended features. After training the model User got to see test results, model accuracy and a few important factors. Users can download the model built by him/her.

# **Literature Review**

Ms Komal[3] published a paper in 2018 that explained big data analytic tools and frameworks. For academics, research, and the IT sector, big data analytics is now essential. The exponentially expanding digital information is primarily in an unstructured form, such as Facebook posts and likes, tweets, blogs, news articles, YouTube videos, website views, etc. It is travelling over the internet infrastructure at a breakneck speed. Through mobile phones, laptops, and PDAs, billions of individuals download, upload, and exchange information on social media and other platforms every day. This study intends to provide in-depth analysis and comparative evaluation of the most recent big data analytics tools and frameworks.

According to the authors Nguyen, Giang & Dlugolinsky, Stefan & Bobak, Martin & Tran, Viet & Lopez Garcia, Alvaro & Heredia, Ignacio & Malík, Peter & Hluchý, Ladislav,[2] the development of software in the field of Machine Learning moves quickly, and a lot of it is open-source software that originates from academic institutions, businesses, startups, or larger open-source communities. This study offers a current time-slide complete overview, comparisons, and trends in the creation and application of cutting-edge AI software. Additionally, it gives a general overview of enormous parallelism support, which may scale computation in the Big Data age effectively and efficiently.

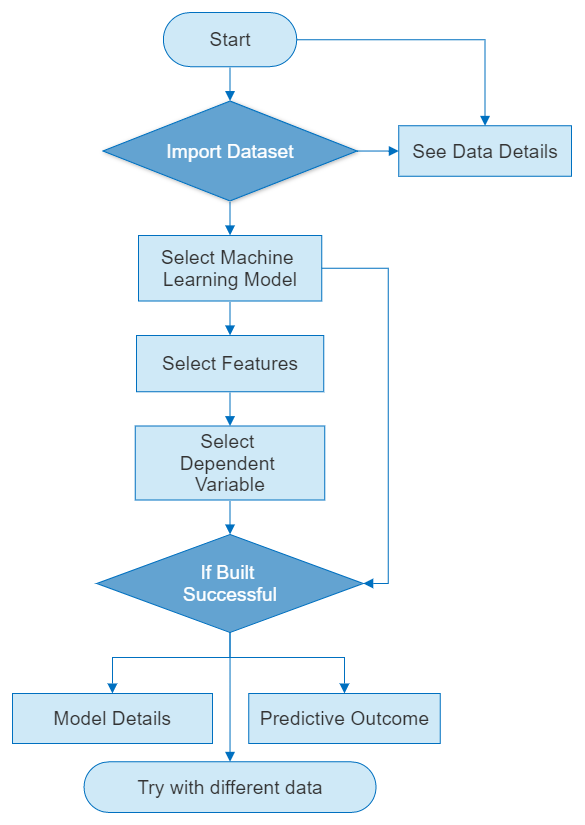
To analyse big data analytics technologies for organisations' advantages, the project's goal was to provide a solution using a model. Kangelani, P. & Iyamu, T. [1] used a case study approach, semi-structured interviews, and a qualitative methodology. Criteria, business and IT alignment, governance, and skill-sets were identified as significant elements of the investigation. A model that may be used as a foundation for evaluating big data analytics technologies within an organisation was created based on the factors.

In addition to offering a classification system that may be used to examine open-source data science software, Barlas, Panagiotis & Lanning, Ivor & Heavey, and Cathal[5] have provided an overview of open-source data science tools. On general features, project activities, operational characteristics, and data mining characteristics, the proposed classification scheme is built. The establishment of a new classification system for analysis and research on open source data science tools is the survey's main contribution. This report also offers a summary of current open-source data science tools.

Big data is typically a huge collection of data that is difficult to manipulate or analyse using conventional data mining or data analytics approaches. Therefore, big data analytics using contemporary big data analytics tools must deal with many difficult tasks such as data gathering, processing, searching, storing, sharing, and privacy as well as data visualization. In this study, Dr A Suresh, T.M. Palayam, Dr R. Kumar, Dr E. Kannan, and Dr Sagunthala [4] have discussed a survey of current big data analytics methods for addressing the aforementioned problems.

# **Methodology/Experimental**

## Algorithm/ Block Diagram



*Fig.1*

**Upload Dataset:** Can use any dataset just the user has to upload the dataset and give appropriate to it.

**View Dataset:** View the dataset easily.

**Plots:** View widely used plots.

**Feature Relations:** Correlations and feature relevance.

**Data Details:**

1. Viewable Dataset itself
2. Data Details
3. Correlation Table
4. Necessary plots

**Select Machine Learning Model**: Select the available machine learning model for uploaded datasets according to need.

**Select Features:** Select the best features recommended (on basis of correlation) by software for the best outcomes of the model.

**Select Target Variable:** Select your target variable or variable that should be predicted.

**Model Details:** View well-trained model details.

**Export Model:** User can export raw code of process he/she did through software in .py/. ipynb/.txt file format

## Technology Used

**Python:** An object-oriented, multifunctional, and high-level programming language.

**Python libraries:**

1. Numpy
2. Pandas
3. Matplotlib
4. Seaborn
5. Sklearn

**HTML:** Website Markup Language

**CSS:** Design Language

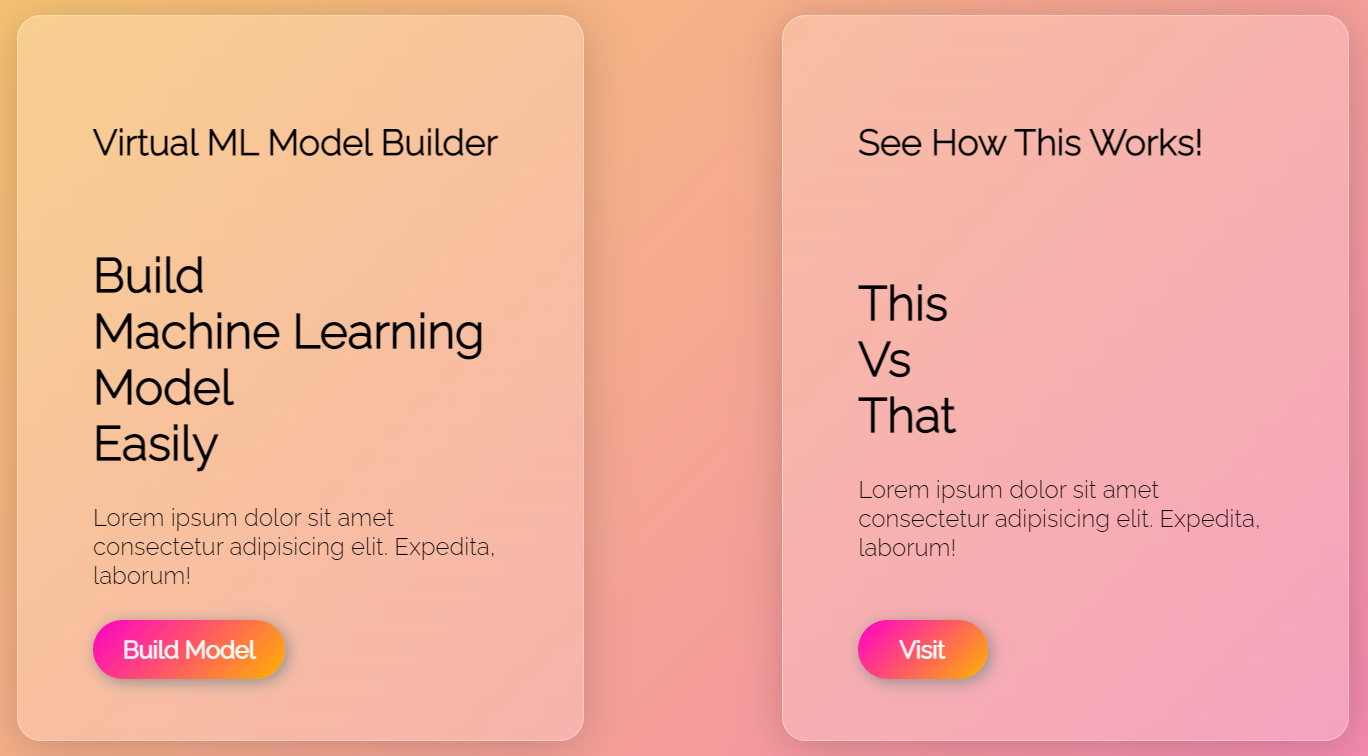
**JavaScript:** Scripting language.

**Django:** A Full-stack framework based on python.

**Database:** MySQL

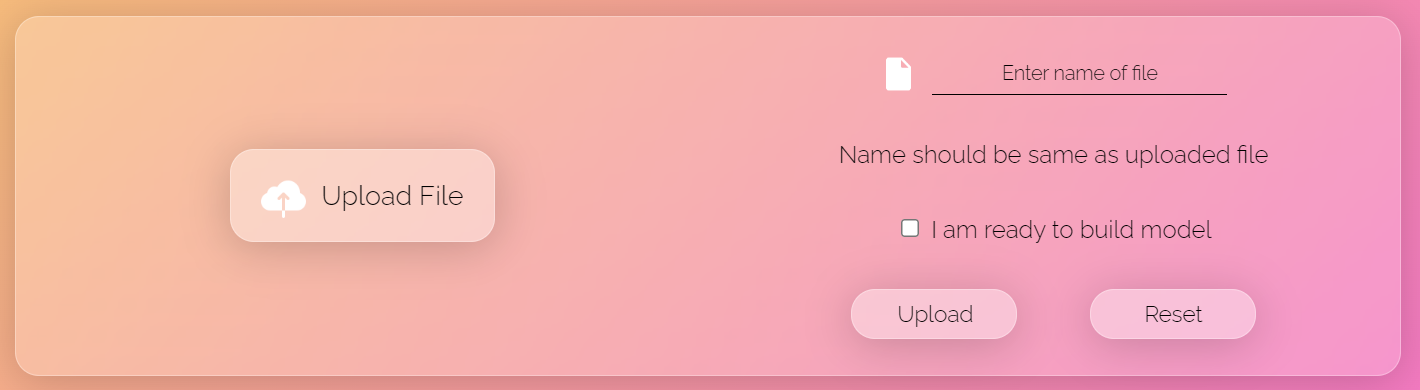
# **Result & Discussions**

For the machine learning model builder, I have used the Django framework and added beauty with CSS.



*Fig.2*

The web application starts with a beautiful UI as Fig.2 shows.

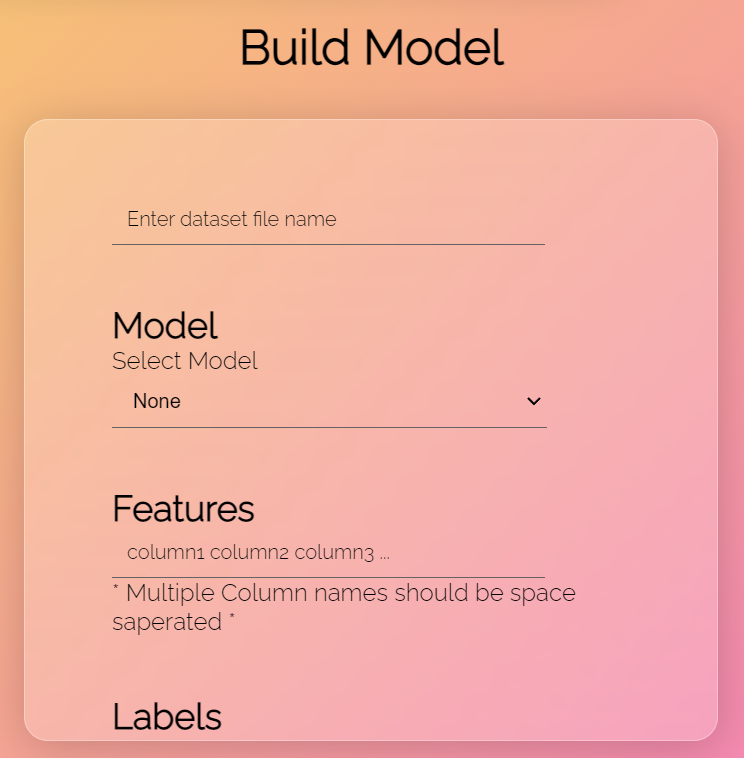


*Fig.3*

When we proceed, it will redirect to the upload file section which looks like this in Fig.3.

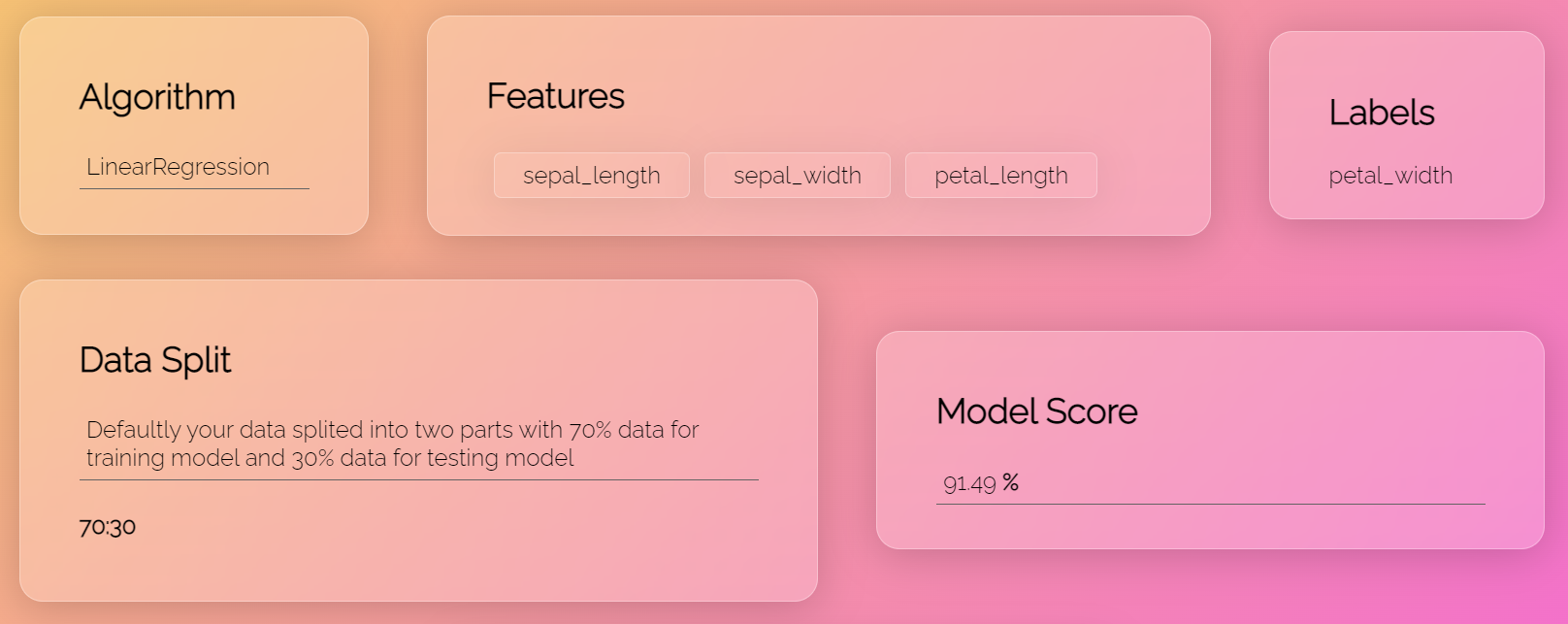


*Fig.4*



*Fig.5*

After uploading the dataset file, it will show little information about the data like in fig.4. This much information is enough for building, training and testing models like in fig.5.



*Fig.6*

After a well-trained model, we will get model details as shown in fig.6.

We can export this model as a text file and build new models with different datasets.

# **Limitations**

This web application provides a limited number of supervised machine learning algorithms. The dataset should contain features that are in numerical format for training the machine learning model. There is a lake of analyzing tools that are required for data preprocessing and visualizations are restricted to a few plots. The project is restricted to only a few models. The project might crash if we provide inappropriate features for the training model.

# **Future Scope**

In future, There Could be datasets imported directly from Kaggle. There could be several options for data preprocessing and exploratory analysis. There could be more available for machine learning models.

# **Conclusion**

The machine learning model builder is a very effective tool for machine learning aspirants and it has all the necessary features that every aspirant needs. Users need not remember the flow or cycle of machine learning, all things work in flow. This project definitely will save users time. It accepts any dataset file of any size with a file extension of "comma-separated variables". Since software provides few analyzing as well as visualizing tools so there is no need to do such stuff. Users can export the code after getting the desired model.

# **Acknowledgement**

We are thankful to our prof. Shital Sobale for guiding and motivating me through the project and supporting me for best results it.

**References**

1. Kangelani, P., Iyamu, T. (2020). A Model for Evaluating Big Data Analytics Tools for Organisation Purposes.
2. Nguyen, Giang & Dlugolinsky, Stefan & Bobak, Martin & Tran, Viet & Lopez Garcia, Alvaro & Heredia, Ignacio & Malík, Peter & Hluchý, Ladislav. (2019). Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey. Artificial Intelligence Review. 52. 77-124. 10.1007/s10462-018-09679-z.
3. Komal, Ms. (2018). A Review Paper on Big Data Analytics Tools.
4. Dr A Suresh, T.M.Palayam, Dr.R. Kumar, Dr.E. Kannan, Dr Sagunthala,(2018). A Study on Modern Big Data Analytics Tools.
5. Barlas, Panagiotis & Lanning, Ivor & Heavey, Cathal. (2015). A Survey of Open Source Data Science Tools. International Journal of Intelligent Computing and Cybernetics, http://www.emeraldinsight.com/doi/abs/10.1108/IJICC-07-2014-0031. 8. 10.1108/IJICC-07-2014-0031.