

# Startup Idea Validation using Machine Learning

Sudhanshu Kumar, Bhargav Guirish, Akshay Umale, Pramod Ganjewar



**Abstract:** According to experts, 90 out of 100 Startups fail within the first year of their launch. There are several reasons behind that but one of the very first and major reasons is Idea Validation. The Entrepreneurs are so excited about their idea that they forget about this very important step, which fails the Startup. Startup Idea Validation tool provides the entrepreneurs with a roadmap by asking the right questions relevant to their ideas and providing suggestions and improvements based on the inputs provided by the User. We are using Linear Regression and Support Vector Machine to train the machine on the Expert Dataset provided and we conclude the result based on that. With this, we'll be able to train the machine with an accuracy of up to 99.99%. There are no other tools available so far which can validate a Startup idea automatically. This tool can be a great breakthrough in the field of Startups.

**Keywords—** Startup; Entrepreneurship; Decision Tree Classifier, Machine Learning, Idea Validation

## I. INTRODUCTION

For almost more than a decade Startup has been a burning topic among the Engineers, Students. Almost every millennial thinks at least once about doing a Startup, maybe for money or for being their boss or to make some difference in the society. But, we know very well that 90 out of 100 Startups fail within the first year of launch. There is not just one reason behind this but there are many factors that play a role behind the success or failure of a Startup. According to one of the articles on cbinsights.com, the number one reason for the failure of a Startup is- **No Market Need** and along with that there are 19 other major reasons they have mentioned like Funding, Team, Competition, Pricing and so on.

With our Startup Idea Validation tool, we are trying to give users an expert opinion on their Ideas and save them from spending a lot of time and energy on the Startup Idea that will not work.

Coming up with Startup ideas can be hard, but validating them is even harder, and more important to your Startup than

you might realize. Because regardless of how world-changing your idea is, if you build a product before validating the idea it's based on, there's a good chance your product will fail. That is where the Idea validation software comes in the picture. Currently, it is being done manually by some industry experts and they charge for it. Also, it takes a lot of time. We will be validating the idea with the help of Artificial Intelligence and Machine Learning. It's a Web-based application that can be accessed from anywhere in the world. All one needs is a computer with an Internet connection. It's a platform-independent application so it can be run on any sort of OS and machine, e.g. Mobile Devices, Desktop Computers, etc.

The rest of the paper is organized as follows. Literature Survey is provided in section II. Problem Definition and Scope are explained in section III. Proposed embedding and extraction algorithms are explained in section IV. Machine Learning Techniques are explained in section V. Implementation steps are given in section VI. Experimental results are presented in section VII. Concluding remarks are given in section VIII.

## II. LITERATURE SURVEY

Ideas are like sparks that glow greatly at the instant they occur, but if they're not maintained, they'll easily shut down. Most people during an instant of our lives had an excellent idea, which we were convinced would change the globe, but only a tiny low number of individuals can make the thought concrete and turn it into a business, without letting the spark shut down.

If you are in the category of a fair amount of people, the great idea that is shining in your mind is exactly the reason why you decided to start your own business. You are ready to separate the valuable resources necessary for your idea to reach out to other people. Before you start investing time and money into your business development, you need to check whether this great idea is great only to you, or is something that can contribute to solving specific problems and make people's lives better. In other words, it's necessary to test and validate your idea. Although the absence of money and inability to execute the idea is often mentioned as reasons for failure, the root cause is usually lack of proper market research. It is very difficult to tell sometimes if it's a good idea or a bad idea. Almost all the ideas look good until the Startup doesn't give the result they expect. There are cases when the problem is actually there but for a very small number of people, in this scenario trying to solve the problem will not create a big Startup; also, such problems are being solved by people themselves by some means already. Startups fail because of many reasons. Sometimes some of the unconventional ideas manage to thrive and some of the greatest ideas manage to fail.

Revised Manuscript Received on May 15, 2020.

\* Correspondence Author

**Sudhanshu Kumar\***, Computer Science and Engineering, MIT Academy of Engineering, Pune, India. Email: mail.sudkumar@gmail.com

**Bhargav Guirish**, Computer Science and Engineering, MIT Academy of Engineering, Pune, India. Email: bhargav.gv@gmail.com

**Akshay Umale**, Computer Science and Engineering, MIT Academy of Engineering, Pune, India. Email: aumale3087@gmail.com

**Dr. Pramod Ganjewar**, Computer Science and Engineering, MIT Academy of Engineering, Pune, India. Email: [pdganjewar@comp.maepune.ac.in](mailto:pdganjewar@comp.maepune.ac.in)

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Let's take an example of Airbnb, a successful Startup of today's time sounded very unrealistic and founders were skeptical about this in the beginning. Let's take another example, Beepi.com, a Startup that managed to secure the funding of 60 million dollars and yet failed. Beepi was a used car selling and buying platform launched exactly at the time when this kind of market was growing and yet managed to fail. The reason behind the failure of this Startup was poor management and execution of the company. It will not be an exaggeration if I emphasize the fact again quoting a heading of an article from forbes.com that 90 out of 100 startups fail.

According to get2growth.com, there are about 300 million startups and about 100 million startups are launched annually. This means that only 10 million startups manage to thrive after launch and 90 million startups fail annually. That's data that changes everything. If we assume that only 5 people are involved in one startup, that gives us  $90 \times 5$  million = 450 million people's dreams are shattered every year just because of poor decision which was driven by making unguided and uninformed judgments.

The startup is need of any economy in the World. A Startup help strengthens the spine by any economy by creating more jobs, proving better facilities, generating finance for the country, etc. Also, it makes many people less dependent on government schemes and facilities. Hence it reduces the burden on the government. People realize its potential and this results in a better lifestyle, stronger economy, higher per-capita-income, and happier societies.

Validating ideas means realizing that if the solution to the problem offered by you is solving the problem and whether it's you and your resources can solve the problem or not. This is the very first step towards building a successful startup of any type. Idea Validation is not domain-specific. Let it be any type of startup or venture, validation is the first thing you want to do if you want your startup to succeed. Validating an Idea removes a lot of ambiguity from your plan. You will be able to see the forthcoming path to the journey. It gives you a clearer vision. It tells you exactly what you need to run this startup or you should even go for this startup or not in the first place. Imagine, you spend all of your and your partners' savings on the startup which is going to fail. In this case, it's not only the money you lost, but it's also the precious time you had and a lot of hard work you did. If we are a little informal here, it'll be like spending money on a girl who never loved you and will never love you.

Idea Validation involves Market research, Demand Analysis, Team Analysis, Plan Assessment, Financial Assessment, Product Assessment, and Analysis of the Pain-point you are trying to solve.

### III. PROBLEM DEFINITION AND SCOPE

Idea Validation currently is done by the experts who have some experience in the field and who knows ups and downs and demands of the field. This process takes a lot of time since sometimes even the experts have to go through the manual process of gathering information from the internet or other relevant sources. Experts also charge a good sum of money to help startups validate their ideas. As a startup founder who most of the times are students, don't have enough money and the reach to get the right person to validate their idea. Also, there can be many ideas that need to validate. To solve this problem and save the founders some time, we are creating a

startup Idea validation tool using machine learning which will provide the founders with the right feedback on which they can work and either stop or continue the venture based on that.

## IV. PROPOSED ALGORITHM

### A. Create Startup Validation Expert Dataset–

We first had to create a dataset of various factors/parameters affecting the success of a Startup. We added more than 50 parameters out of which 23 are common for all the Startup types and remaining varies based on the types of a Startup.

Below is the list of parameters on which we are assessing a startup Idea--

1	Common Parameters		
a.	Demand for product	Service or Product	App-based or physical product
b.	Type of Location	Price of the Product	The uniqueness of the Product
c.	Size of Customer-base	B2B/ B2C	Use Frequency
d.	Customer Persona	Founder's experience	Founder's skills
e.	Team	New Features	The requirement of new Features
f.	Required Capital	Capital Source	Customer Survey
g.	Money Required for survival for a Founder	Break-even Point	Profit per unit
h.	Estimated Number of Units sold per year	Revenue Streams	

2	Sector Specific Parameters		
	E-commerce	Marketing	Electric Vehicles
a.	Delivery Speed	Digital or Offline?	Availability
b.	Mobile App Standard	SEO	Ease of Use
c.	Support	SMO	Strength
d.	Quality of Product	Keyword Research	Top Speed
e.	Type of Services Offered	Content Quality Suggestions	Safety Measures
f.	Delivery Options	Location of Ad	Mileage of Single Charge
g.	Availability of product	Cost of Creating Campaign	Load Capacity
h.			Charging Facility

2	Sector Specific Parameters		
	E-commerce	Marketing	Electric Vehicles
i.			Maintenance Facility

As of now, we are using only three types of Startups which are **marketing, E-commerce, and Electric Vehicles**. We can add as many types as possible by constantly adding up the data in the dataset and running the algorithms

It's very important to check for a large number of parameters to be sure about the result. Although, the result in the datasets are created based on the certain conditions provided by the experts and the outcome depends totally on the information provided by a user.

**B. Encoding**

In our project, we are using OneHotEncoder to represent categorical variables as a binary vector. To do that, categorical values are mapped to Integer values. Then mapped integer values are represented as a binary vector is all valued zero except the index of Integers. The Index of integers is marked with a 1.

It makes the categorical data more expressive. Not all machine learning algorithms can work with categorical data directly. Its conversion from Categories to Numbers must be done. Both Inputs and Output variables are required to be encoded if they are categorical.

We could utilize a whole number encoding straightforwardly, rescaled where required. This may work for issues where there is a characteristic ordinal connection between the classes, and thusly the whole number qualities, for example, names for temperature 'cold', 'warm', and 'hot'.

There might be issues when there is no ordinal relationship and permitting the portrayal to incline toward any such relationship may be harming to figuring out how to take care of the issue. A model may be the marks 'pooch' and 'feline' In these cases, we might want to give the system progressively expressive capacity to become familiar with a likelihood like a number for every conceivable name esteem. This can help in making the issue simpler for the system to display. At the point when a one-hot encoding is utilized for the yield variable, it might offer a more nuanced set of expectations than a solitary name.

**C. Support Vector Regression(SVR)**

SVM is different than any other Regression Models. To predict continuous variables it uses Support Vector Machine, a Classification Algorithm. Linear Regression Models try to minimize the error between the predicted and actual value but in SVR, it tries to fit the best line between a threshold or a fixed error value. It tries to classify.

All the prediction lines are classified into two types. One passes through error boundary and others do not. Lines that do not pass the error boundary are not assumed as the difference between the predicted value and the actual value. The potential support vector is those lines that pass. In our project, we are using this as an alternative to Linear Regression.

**D. Linear Regression**

Linear regression is one of the oldest methods. People say it goes back to 200 years. Linear regression has been

studied from every possible angle and each angle has its different name.

It's a linear model. This means it assumes a linear relationship between the input variables which are X and the single output variable which is Y. In clear language, the value of Y can be calculated from the linear combination of the input variables (X).

It's is referred to as **Simple Linear Regression** if there is a single variable(x) and is called Multiple Linear Regression if there are multiple numbers of Input variables.

Although several techniques can be used to train the linear regression equation from data the most common one is Ordinary Least Squares.

▪ *Linear Regression Model Representation*

This is represented as a linear equation that includes a specific set of Input values (X) and the solution to which is a predicted output (Y).

One scale factor is assigned to each input value or column by a linear equation, it's called Coefficient and is represented by capital Greek Letter β. To give the line an extra degree of freedom, an additional coefficient is also added which is often called the intercept or the bias coefficient.

A simple representation of the regression model would be-

$$Y = \beta X + C$$

The line is called a plane or a hyper-plane in the higher dimensions where we have more than one input (X). Therefore the representation would be of the form of the equation and the specific values used for the coefficients.

It removes the influence of the input variable on the model when the coefficient becomes zero and therefore from the prediction made from the model.

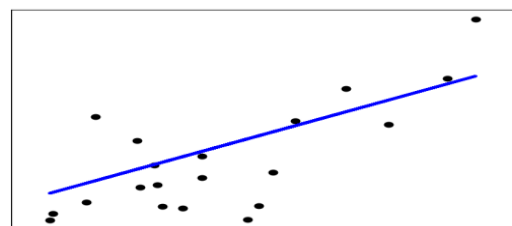
▪ *Making predictions with Linear Regression*

Making a prediction is equally as simple as solving the equation for a certain set of inputs.

Let's take an example here. Let's say we want to predict weight (y) from height (x). So, the linear regression model representation of this problem should be –

$$y = B_0 + B_1 * x_1$$

Here, B<sub>0</sub> is the coefficient of bias, and B<sub>1</sub> is the coefficient for the height and column. We find a good set of coefficient values using a learning technique. Once the machine learns we can provide a different set of values of height to predict the weight.



**Figure 1: Linear Regression Plotting example**

▪ *Linear Regression working*

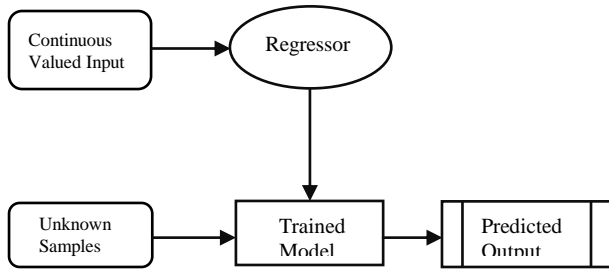


Figure 2: Working Diagram of Linear Regression

A set of input variables are fed to the Regressor for training. Once the model is trained we input unknown samples also called test inputs to the Trained model and we get the required predicted output.

V. IMPLEMENTATION

- A. We imported all the libraries required to train the model. Some of the major libraries are Sklearn, Pandas, Numpy, Linear Regression, etc.
- B. After that, we imported the dataset we had created. The dataset contains more than 150000 rows and 21 columns.
- C. Train the model

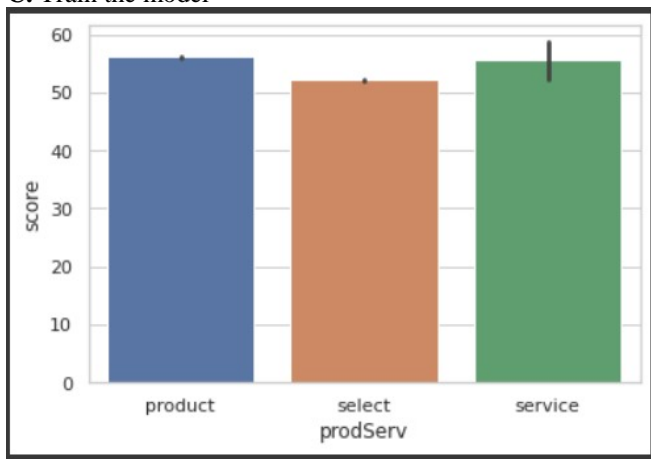


Figure 3: Product/Service Vs Score Trained Model

- D. Create a pickle file to be used in the Django Framework
- E. Import the Pickle file in Django and input the test values to get the machine to predict

VI. EXPERIMENT AND RESULT

The Dataset that we have created gives a float value as an output which is the score defined based on certain conditions. We are using that score to determine the Success or Failure of the Startup and we called it Grade. Also, based on the input provided by the user we provide two different feedbacks. The first feedback is the Observations which are the positive aspects of the startup and the second feedback is Suggestions which is the improvement that we suggest based on the input provided.

Let’s take an example of a set of data here –

Sr. No.	Test Data
---------	-----------

	Variable	Value	Value
1.	Product or Service	service	99.62
2.	App	Yes	
3.	Location	Tier 2	
4.	Price	Same	
5.	Unique	Yes	
6.	Cust. Size	Large	
7.	selling	B2B	
8.	useFrequency	Daily	
9.	Demand	Yes	
10.	Persona	Type1	
11.	Experience	Yes	
12.	havePeople	Yes	
13.	haveSkills	No	
14.	newFeatures	yes	
15.	RequriedNewFeatures	Yes	
16.	haveRequiredMoney	Yes	
17.	didCustomerSurvey	Yes	
18.	AppStandard	Latest	
19.	delivery speed	1 Day	
20.	Support	Yes	
21.	funding sources	available	

VII. CONCLUSION AND FUTURE WORK

With a good dataset and with the help of Machine Learning we can now predict the future of a Startup and save hundreds of entrepreneurs from making the common mistakes which will improve the rate of success of Startups. This will also save the entrepreneurs from spending a lot of money on the ideas which are not even necessary. As they say, if you fail to plan, you plan to fail and here we are giving entrepreneurs the tool which will help them plan their journey carefully.

We’ll further try to add more types of startups and also automatically update the datasets to provide users with the most recent and updated information and feedbacks.

REFERENCES

1. Gulden Kaya Uyanik, “A Study on multiple linear regression analysis”, 4<sup>th</sup> international conference on New Horizon in Education
2. Diti Modi, Linear Regression Model, <https://towardsdatascience.com/linear-regression-model-899558ba0fc4>
3. Jason Brownlee, Linear Regression for Machine Learning, <https://machinelearningmastery.com/linear-regression-for-machine-learning/>
4. Benjamin Jones, Linear Regression Projects, <https://www.kaggle.com/bdmj12/linear-regression-portfolio-project>
5. Omair Aasim, Predict Salary using simple linear Regression, <https://medium.com/analytics-vidhya/machine-learning-project-1-predict-salary-using-simple-linear-regression-d83c498d4e05>
6. C.B insights, Top 20 Reasons Startups Fails, Article on cbinsights.com
7. Chirag Sehra, Decision Tree Explain Easily, Blog on Medium.com
8. Julia Kailyanen, Idea Validation: Steps and Tools for Testing YourIdea, viima.com
9. Max Babych, What is Idea Validation and Why It is a must-do step for any Startup founder, spdload.com
10. Jim Semick, Lean Market Validation, productplan.com
11. Jacob Firuta, 5 Types of Customers and how to approach them, Livechat.com, <https://www.livechat.com/success/types-of-customers/>



12. Tom Treanor, 15 Questions to validate your startup business idea,medium.com, <https://medium.com/@rtmixmktg/15-questions-to-validate-your-start-up-business-idea-d445d192a995>
13. Sujan Patel, 20 Questions you can ask to validate your startup idea, entrepreneur.com, <https://www.entrepreneur.com/article/241122>