

# Career Guidance System Using ML

Vaishnavi S. Kulkarni<sup>1</sup> Amruta S. Khalate<sup>2</sup> Jaydeep H. Natakale<sup>3</sup> Poonam C. Bhosale<sup>4</sup>  
 Dr. Reshma Atole<sup>5</sup>

Student

<sup>1,2,3,4,5</sup>Department of Computer Engineering

<sup>1,2,3,4,5</sup>SVPM's College of Engineering Malegaon(bk), Baramati, India

**Abstract** — Selection of appropriate career selection is one the key factor in students' life. A wrongly chosen career sometimes destroy students' career. Right career selection for a student is always based on selection of right educational program. Career choice is based on the understudy execution and their interface. Choice of fitting instructive program needs a recommendation system. This project is recommending suitable career way according to students' performance. It will save time, efforts. To achieve this objective, suggestion of understudy execution and recognizable proof of career way framework is working.

**Keywords:** Lasso Linear Regression, Influence Factor, Career Guidance, Machine Learning

## I. INTRODUCTION

The focus of this project is to analyze student performance based on skills or other basic activities and recommend a career path for student and it is more useful and important. To be able to recommend the student performance using the ML algorithm. Choosing of educational program needs a recommendation system to for students' performance prior to make career in specific field. To determine student performance that improve students' knowledge, performance and career path of specific student. In Today's world, many students are facing the problem how to choose their career path. Students quit confuse. So, it's help to choose career in better way. Those strategies can be planned after analyzing the student's performance, since the advanced estimation of failure rate can help education.

The main objective of the project is to predict student performance based on influence factor and display the graph and also gives the career path for that student. To be able to make the performance of student more efficient for them. This project is able to predict the student performance using the ML algorithm. In Today's world many students are facing the problem how to choose their career path. Students quit confuse. As students are refer through their academics and pursuing their interested courses, it is very important for them to access their capabilities and identify their interests.

One of the most challenging tasks in the education sector is to predict students academic performance due to huge amount of student data. This is a very big problem to arbitrate many characteristics related to students. These characteristics can be divided into students marks, skills, grade, certifications, interests and hobbies.

1) The commonly used data mining techniques to predict student performance in previous studies is to find out the most suitable technology. There are several technique to predict student performance .Lasso regression is accurate in predicting the exact final grade of students. Machine learning technique in educational data mining aim to develop a model.

- 2) A study on the prediction of student performance by Angeline D M ,it was used apriori algorithm in which it extracts the set of rules specific to every factor and analyze the given performance to classify assignment, internal assessment, group action etc. In this it was divide the given set into average, good, below average.
- 3) Literature survey done using Hybrid Regression and Multi-Label classification, the Abdullah alshanqiti and abdallah namoun used the different influential factors for predicting student performance. They used both supervised learning and unsupervised learning algorithm to predict student performance. In this survey they identify student failure rate was considered. K-NN, self-mapping and NN these concepts were used.
- 4) Prediction of student performance using Machine learning and data mining application for improving the learning experience of students by early prediction of student performance can support this process by detecting the low performance of the student and it will help students to improve their educational performance and keep the student motivaed.
- 5) After researching many papers, we will use Lasso Linear regression to predict student performance and for input we consider some factors in binary format, fuzzy set etc.

The purpose of this is to predict student performance based on the factors such as marks, skills, interest and suggest a career path for student and it is more efficient and accurate. This paper mainly describes the prediction of students performance and at last suggest a career path of particular student.

## II. SYSTEM ARCHITECTURE

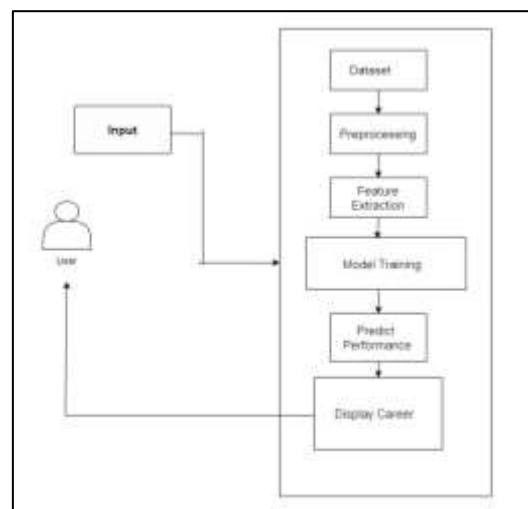


Fig. 1: System\_Architecture

### III. MODELLING AND ANALYSIS

#### A. Lasso Regression

We consider Lasso regression model, an optimized type of direct regression. Gradationally from the facticity that our dataset X is entirely formalized, it describes nonstop features exactly and doesn't clench any misplaced evaluations. hidden, the incommodity of having impracticable features or data outliers subsists. accordingly, the use of Lasso model can upgrade the overall feast of m3in two fetches; it simplifies the model by executing a juncture alternative latently, which neglects the nonpractical features that have no significant authority; it estimates hard up futures whilst adjoining a despite as are regularizer, which prevents the miniature from being overfitted. More just, the entity of Lasso is to minimize.

$$\underset{w}{\operatorname{argmin}} \left\{ \frac{1}{2k} \|Xw - g\|_2^2 + \lambda \|w\|_1 \right\} \dots(1)$$

X = Dataset of pupil using regression model

W = means vector

G = Grade  $\| \cdot \|_2$  is 2- normal, and  $\lambda$  is a unchanging that is generally set to 0.1.

Once the smart measure vector w is set up, we can calculate the prophecy for a given input X using a simple direct regression model described

$$m^3(x_i) = w_0 + \sum_{n=1}^j w_n(x_{i,n}) \dots(2)$$

where w0 is the intercept.

XGBoost denotes extreme Gradient Boosting.

Alpha value is = 1.0

Coefficients vectors w, h

W is denotes as w info\_screenwidth

H is imported as w info\_screenheight In our system we load student dataset from scikit learn dataset module

We resolve data into predictors(X) And response variable(y) and resolve the data into training and testing sets.

Next we initialize a lasso regression module with regularization parameter of 0.1 and befit to practice data using the fit system. We take predictors on test data using prognosticate system and calculate the mean squared error function.

From sklearn.metrics module eventually we publish measure of model using a pandas dataframe which use relationship of point and response variable.

The accuracy of lasso regression is 9.123823316437365

#### B. Multilabel Classification

To incorporate multilabel classification in a student career recommendation and performance system using machine learning, you can follow these general steps:

- 1) **Data Preparation:** Collect and preprocess the data related to student information, academic performance, and career options. Ensure that the data is labeled with multiple categories or labels representing different career paths.
- 2) **Feature Engineering:** Extract relevant features from the available data that can contribute to predicting career recommendations and performance. This may include

factors such as academic grades, extracurricular activities, internships, skills, interests, etc.

- 3) **Data Split:** Divide the dataset into training, validation, and testing sets. The training set will be used to train the model, the validation set will be used to fine-tune the model and optimize hyperparameters, and the testing set will be used to evaluate the model's performance.
- 4) **Model Selection:** Choose an appropriate multilabel classification algorithm for your task. Some common algorithms for multilabel classification.
- 5) **Model Training:** Train the selected model on the training data. The model should learn to predict multiple labels simultaneously based on the provided features.
- 6) **Model Evaluation:** Evaluate the performance of the trained model using the validation set. Common evaluation metrics for multilabel classification include precision, recall, F1-score, and accuracy.
- 7) **Hyperparameter Tuning:** Adjust the hyperparameters of the model to improve its performance. This can be done using techniques like grid search or random search.
- 8) **Model Deployment:** Once you are satisfied with the model's performance, deploy it in your student career recommendation and performance system. This will involve integrating the model into the system's architecture and making it accessible to generate career recommendations based on input student data.
- 9) **Testing and Monitoring:** Continuously monitor the system's performance and gather feedback from users. Regularly test the model's accuracy and update it as needed to improve its recommendations and predictions.

#### C. Multicollinearity

Multicollinearity refers to the sight of high correlations between predictor variables (features) exploited in the Lasso reversion model. Multicollinearity can pose objections in the model's reading and illustration. Lasso reversion is a regularization form that combines attribute choosing and regularization by charging a despite on the correct worths of the reversion amounts. It encourages light answers by shrinking the many of inapplicable features to zero. even so, when multicollinearity exists, Lasso reversion may front hardships in taking the most material features or delivering strong amount evaluations.

#### D. XGBoost:

It works on gradient boosting algorithm Gradient boosting algorithm workshop on the introductory principle gradient descent. This miniature is made up using tree- predicated learners( Decision Trees) Boosting works on the principle of ensemble ways where violations from earlier models are reduced by the new models .

XGradient boosting Algorithm:

Final prediction = Base value( the starting prediction from introductory decision tree)

LR \* w1 LR \* w2. LR \* wn

Where LR = education rate = eta

w1 = residual predicted value by 1st residual model

wn = residual predicted value by utmost residual model

Xgboost is dissimilar from other gradient boost is because of its tuning parameters

The main tuning parameters are:

- 1) regularisation parameter( Lambda)
- 2) threshold that defines machine pruning( Gamma)
- 3) Learning rate( eta)

#### IV. RESULT AND DISCUSSION

In this project, we implemented the career path for the student. Using the student influence factor displaying the performance and career path.

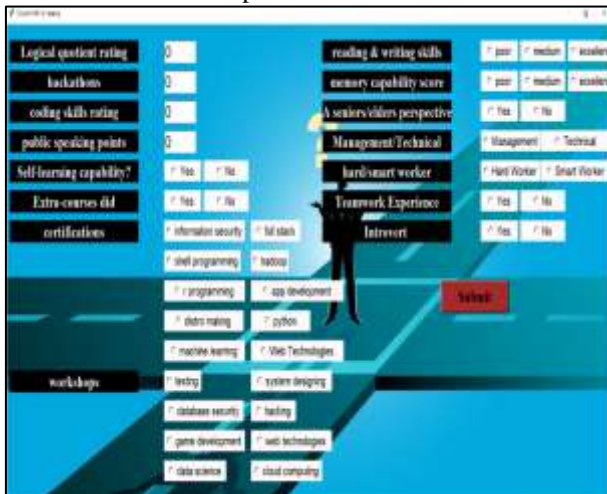


Fig. 1: Predicting performance form

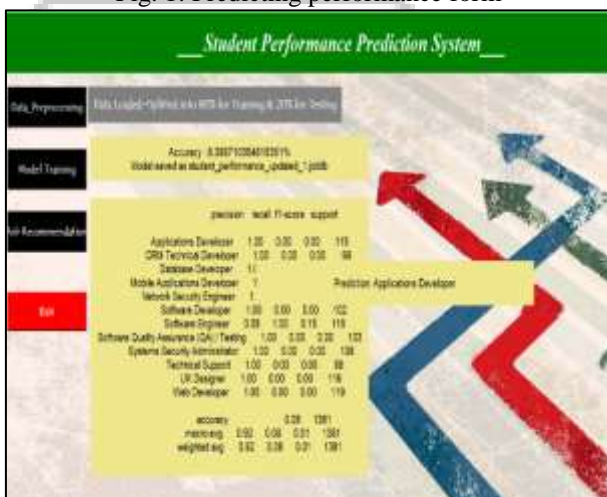


Fig. 2: Career suggestion

#### V. CONCLUSION

In this way, it'll be largely useful in helping scholars learning times. It'll offer the performance of particular scholar applying some agent and it'll helps to detect out the interest of scholar. Using the machine learning algorithm like LASSO which gives the scholar performance using scholar academic record and so on. Also for suggesting the career path for scholar the machine learning algorithm is used. This process can helps to decide easy about performance of the scholars and indicate career path. So, it'll helps to lead the scholar for their career.

#### REFERENCES

[1] Ramesh, V., Parkavi, P., Ramar, K. (2013). Predicting student performance: a statistical and data mining

approach. International journal of computer applications, 63(8), 35-39.

[2] Angeline, D. M. D. (2013). Association rule generation for student performance analysis using apriori al Transactions on Computer Science Engineering its Applications (CSEA), 1(1), 12-16algorithm. The SIJ.

[3] Chalaris, M., Gritzalis, S., Maragoudakis, M., Sgouropoulou, C., Tsolakidis, A. (2014). Improving quality of educational processes providing new knowledge using data mining techniques. ProcediaSocial and Behavioral Sciences, 147, 390-397.

[4] Pauziah, M.A., Arsad, N.B., Jamalul-lail, A. M. (2013, November). A neural network students' performance prediction model (NNSPPM). In Smart Instrumentation, Measurement and Application (ICSIMA), 2013 IEEE International Conference on (pp. 1-5). IEEE.

[5] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Mining multi-label data," in Data Mining and Knowledge Discovery Handbook. Boston, MA, USA: Springer, 2010, pp. 667–685.

[6] T. Kohonen, "The self-organizing map," Proc. IEEE, vol. 78, no. 9, pp. 1464–1480, Sep. 1990.

[7] J. V. G. Tsoumakas and E. Spyromitros. (2020). MULAN: A Java Library for Multi-Label Learning. [Online]. Available: <http://mulan.sourceforge.net/datasets-mlc.html>

[8] G. D. Kuh, S. O. Ikenberry, N. A. Jankowski, T. R. Cain, P. T. Ewell, P. Hutchings, and J. Kinzie, Using Evidence of Student Learning to Improve Higher Education. Hoboken, NJ, USA: Wiley, 2015.

[9] A. Namoun, A. Taleb, and M. Benaida, "An expert comparison of accreditation support tools for the undergraduate computing programs," Int. J. Adv. Comput. Sci. Appl., vol. 9, no. 9, pp. 371–384, 2018.