"A Real Time Application to Predict Student Placement Status"

MOHAMMED SHAMAIL FARHAN

Department of Computer Science and Engineering,

JSS Science and Technology University,

Mysuru , Karnataka ,India

Abstract

Students placements are important at an educational institution, and they can affect the college's reputation. One of the most significant difficulties that higher education institutions confront today is improving student placement performance. When educational entities get more complex, placement prediction becomes more difficult. Educational institutions seek more efficient technology to aid in improved administration and decision-making procedures, as well as to help them develop new strategies. Providing new knowledge about innovative ideas to the management is one of the most effective approaches to solve the issues of enhancing quality. Knowledge may be derived from operational and historical data stored in the databases of educational organisations using machine learning techniques. The data set for system implementation includes student data from the past. These data are utilized to train the model for rule detection and classification, as well as to test the model.

This method proposes a recommendation mechanism that forecasts the placement status of students. This strategy aids an organization's placement cell in identifying potential students, as well as paying attention to and improving their technical and interpersonal abilities. As a result, they will be able to put in more effort in order to get placed in higher-ranking companies.

This method presents a suggestion mechanism for predicting students' placement status. For student placement prediction, data science techniques are used, and the system employs the most efficient algorithms, such as the "naive bayes algorithm" and the "KNN algorithm." Both algorithms are implemented in the C# programming language, and the results are compared, and the system determines which algorithm is the more efficient. Our system produced excellent results, with around 85% for the "naive bayes algorithm" and around 95% for the "KNN method." When compared to the naive bayes algorithm, the KNN approach produces better results.

Introduction

Students who enrol in professional courses at higher education institutions hope to land a well-paying job with a reputable company. Professional education can be entirely technical or include both technical and managerial components. Students in CS Department, E and C Department, Civil Department, Mechanical Department, and other subjects receive technical education through technical course programmes. This degree is designed to prepare students to be specialists in state-of-the-art conjectural and practical knowledge in a variety of engineering fields. The ability to estimate the placement status that degree students are most likely to attain will encourage students to work harder in order to make proper progress toward a career in diverse technical sectors. It will also assist professors and the placement cell in an institution in providing proper care for students' improvement throughout the course. A high placement rate is an important factor in establishing an educational institution's reputation. As a result, any higher learning institution's educational system should include such a system.

Algorithms involving machine learning are typically based on mathematics and statistics. Because it is difficult to handle every condition on an item, the key advantage of machine learning over traditional software is that there is no written code that teaches the system how to make the decision to choose

the proper object among-st two distinct ones. That's where machine learning comes in. It mostly employs the provided data to make intelligent decisions, forecast the future, or find anomalies. It is now utilized in a variety of applications like as self-driving cars, virtual assistants, search engine results, object recognition, advertising, and predictive analysis, among others. Supervised Learning, Unsupervised Learning, and Reinforcement Learning are the three sub-fields that make up machine learning techniques. We apply supervised learning techniques to predict student placement in this proposed study. The proposed work is created as a real-time application for educational institutions. The proposed system was built with Visual Studio as the front end and SQL Server as the back end. For project development, we employ Microsoft technologies because they cover all of the tools, frameworks, and concepts needed to deal with real-time applications. C# is the programming language we employ, and the machine learning techniques we utilise are written in C# because it is more compatible with Visual Studio. We built our own library and utilised it to predict placement.

Related Works

In educational sector , placements pays a vital role and all colleges wants a good results in placements. 100% placements is the reputation of the college. Parents prefers the colleges with a very good placement results and training and teaching. There a so many

previous works related to the placements prediction. Most of the research works are innovative and uses data science algorithms for placements prediction. Even though many works done on students placements prediction, all of the previous works used ready libraries and ready tools for students placement prediction. Most of the works uses "python" as programming language and which supports all ready libraries and packages for machine learning algorithms.

Some of the previous works uses ready tools for students placements prediction , tools such as "WEKA" tool, "Rapid Miner" tool , R Tool etc. Using all these tools students placements prediction is done. All of the previous research works just used the educational datasets and placements prediction is done. None of the previous works developed this as real time application useful for the colleges and educational sector. Most of the previous works uses small datasets and less number of parameters for the students placements prediction.

Gap Analysis

1.In the ieee paper they are predicting only type of the company such as dream company, core company, mass recruiters, not eligible or not interested. But in the proposed system we predict the actual company name using ML techniques.

2.In the ieee paper they have presented an idea of placement prediction, But in the proposed system we build this concept as real time application useful for education sectors.

3.In all ieee papers, algorithms used are not programmed, they have used ready libraries for algorithms and tools used for algorithms. But in the proposed system we program the algorithm means we write our own logic for the algorithm and results will be tested.

4.In ieee papers only few parameters used for prediction, But in the proposed system we use more parameters such as student's attendance, communication skills, grasping capability, technical skills, aptitude, previous semester results, time management etc,....

5.we use front end technology as "visual Studio" and back end technology as "SQL Server" and C# as programming language. visual studio and SQL server are from Microsoft technologies, its supports all concepts and libraries required to work with real time application. C# is more supportive with visual studio. we can build application in less time using these technologies.

3. Proposed Work

- System is an real time application for student placement prediction and can be accessed from different locations.

- System is an educational based application which helps institutes to know the percentage of placements.

- Proposed system clusters the students based on the characteristics or parameters, here we mainly concentrates on knowledge, skills and attitude of the students.

- System predicts the student chances of placements using classification rules technique.

- This system presents a recommendation system that predicts the students to have one of the placement statuses, viz., Infosys, Wipro, Cisco, Accenture etc...

- System makes use of previous data to predict the future.

3.1 Parameters used

- **CGPA**

- **SSLC**

- **PUC**

- **TS**

- **CC**

- **CS**

- **CA**

4.Proposed Methodology

Machine Learning

Machine learning is a data-driven method of examining a system. Machine learning is a

branch of data science in which data is processed using machine learning algorithms.

Supervised Learning Technique

Basically classification is used to classify each item in a set of data into one of the predefined set of classes or groups.

Classification methods make use of mathematical techniques for problem solving

Ex: Employee statuses in a company (leaves or stay)

To predict which current employees are probably to leave in the future..

In this project for student placement prediction we make use to "*Bayesian Classifier and KNN algorithm*". These 2 algorithms generates good results for educational datasets and most of the previous works used these algorithms and got good results, but they have used ready libraries and packages for prediction, but here we write our own logic and student placement prediction is done.

Flow of the Algorithm
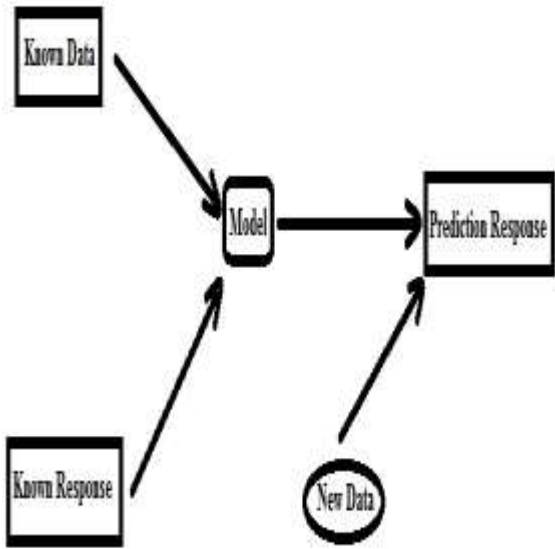
Proposed Model
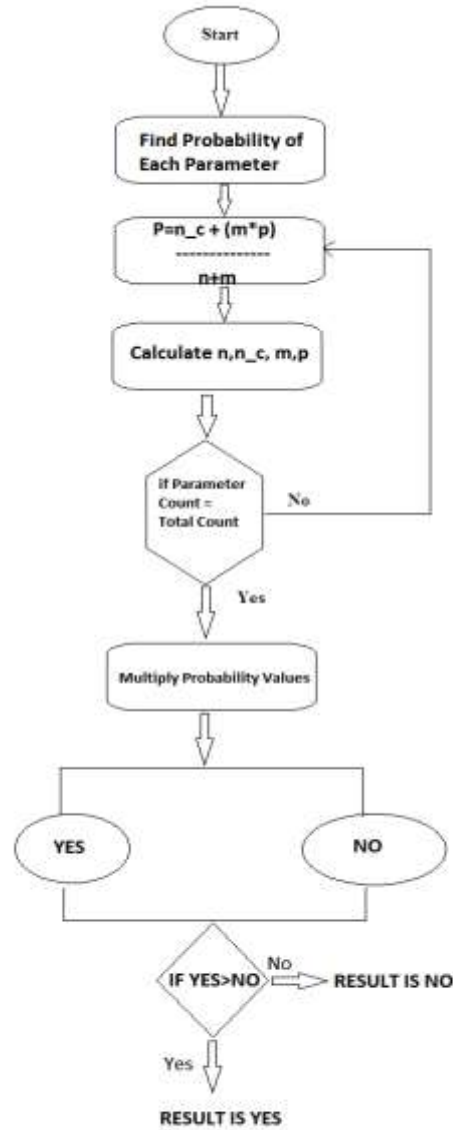
FLOW OF NAIVE BAYES ALGORITHM

4.1 Placements Prediction Process

4.1.1 Bayesian Classifier Algorithm

Bayesian classifier is one of the efficient and good classifier, This classifier works based on the probability calculation which generates high accurate results for educational datasets. In our work we have written the algorithm program using C# our own logic and program works fine for dynamic data.

Student Placement Predictor - Working of Naive Bayes with example

Inputs

1. Parameters - Grasping Capability(P1), Communication Skills(P2), Attendance(P3)

2. Outcome - Infosys, Wipro [p=1/2=0.5)

3. Training Datasets

Students	P1	P2	P3	Result
Anil	Low	Low	Regular	Infosys
Kumar	Low	Average	Irregular	Infosys
Ajay	Average	Average	Regular	Wipro
Sunil	High	Low	Irregular	Infosys
Punith	High	High	Regular	Wipro

Infosys	Wipro
Grasping Capability - High	Grasping Capability - High
$P=[n\_c+(m*p)]/(n+m)$	$P=[n\_c+(m*p)]/(n+m)$
n=2, n_c=1,m=3,p=0.5	n=2, n_c=1,m=3,p=0.5
p=[1+(3*0.5)]/(2+3)	p=[1+(3*0.5)]/(2+3)
p=0.5	p=0.5
Communication Skills(P2) - Average	Communication Skills(P2) - Average
$P=[n\_c+(m*p)]/(n+m)$	$P=[n\_c+(m*p)]/(n+m)$
n=2, n_c=1,m=3,p=0.5	n=2, n_c=1,m=3,p=0.5
p=[2+(3*0.5)]/(2+3)	p=[2+(3*0.5)]/(2+3)
p=0.5	p=0.5
Attendance(P3) - Regular	Attendance(P3) - Regular
$P=[n\_c+(m*p)]/(n+m)$	$P=[n\_c+(m*p)]/(n+m)$
n=3, n_c=1,m=3,p=0.5	n=3, n_c=2,m=3,p=0.5
p=[1+(3*0.5)]/(3+3)	p=[2+(3*0.5)]/(3+3)
p=0.4	p=0.58

4. New Student Classification

Farhan -

Grasping Capability(P1) - High

Communication Skills(P2) - Average

Attendance(P3) - Regular

Placement Status - ?

Algorithm works based on the probability calculation : $P=[n\_c + (m*p)]/(n+m)$

Find probability for all 3 parameters for both Infosys and Wipro .

Infosys – 0.5 * 0.5 * 0.4 * 0.5 (p)

=0.05

Wipro – 0.5 * 0.5 * 0.58 * 0.5 (p)

=0.07

Since 0.07 > 0.05
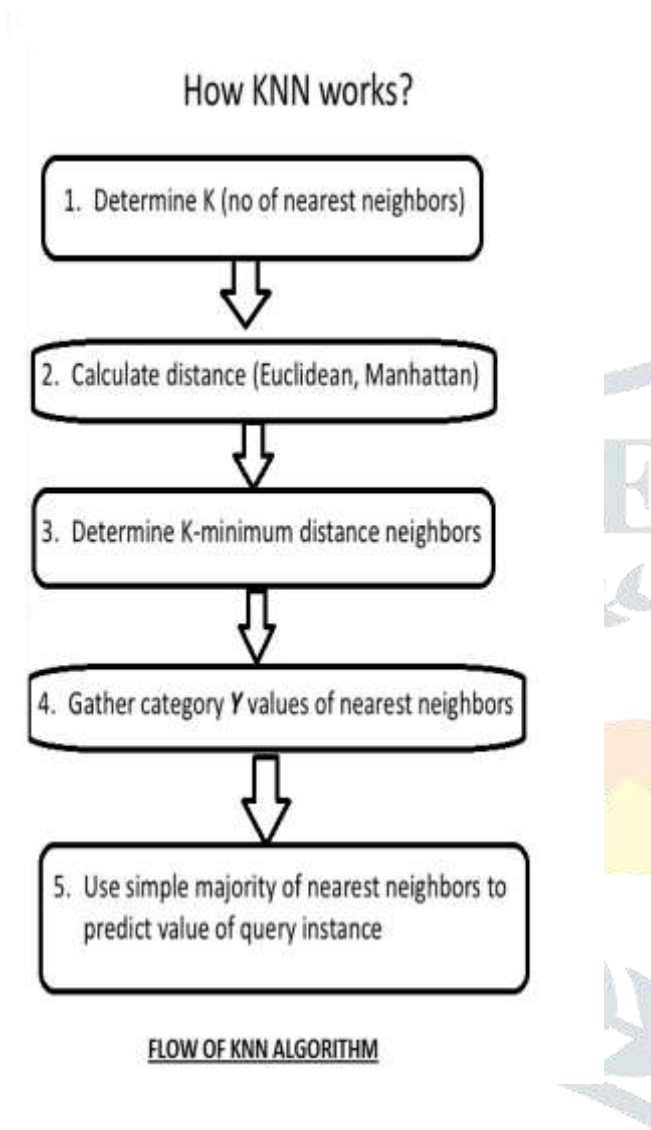
Farhan is classified to **Wipro**

4.1.2 KNN Algorithm Steps

K nearest neighbour algorithm is also one of the efficient algorithm which works only for numbers. It works for numerical data. And process data faster compared to other

algorithms. It works based on the distance calculation.

Flow of the Algorithm

FLOW OF KNN ALGORITHM

Sample Example (Sample Data-set) using KNN technique

Sample Training Datasets

KNN Example

Records	P1	P2	P3	Result
1	7	2	1	+
2	3	5	2	-
3	2	6	2	+
4	6	3	1	+
5	4	4	2	-
6	7	2	1	-
7	2	3	4	+
8	4	3	3	?

1. Determine K

1. Determine value of K

Suppose K = 3

2. Calculate distance

Coordinates of query instance are (4,3,3)

Coordinates of training instance(1) are (7,2,1)

$D = SQRT ((7-4)^2+(2-3)^2+(1-3)^2) = 3.74165$

2. Calculate distance

Records	P1	P2	P3	Result	distance
1	7	2	1	+	3.741657
2	3	5	2	-	2.44949
3	2	6	2	+	3.741657
4	6	3	1	+	2.828427
5	4	4	2	-	1.414214
6	7	2	1	-	3.741657
7	2	3	4	+	2.236068
8	4	3	3	?	

3. Determine K-minimum distance neighbors

K = 3

Records	P1	P2	P3	Result	distance
1	7	2	1	+	3.741657
2	3	5	2	-	2.44949 (3)
3	2	6	2	+	3.741657
4	6	3	1	+	2.828427
5	4	4	2	-	1.414214 (1)
6	7	2	1	-	3.741657
7	2	3	4	+	2.236068 (2)
8	4	3	3	?	

4. Gather category Y values of nearest neighbors

Records	P1	P2	P3	Result
2	3	5	2	-
5	4	4	2	-
7	2	3	4	+
8	4	3	3	

5. Use simple majority of nearest neighbors to predict value of query instance

Records	P1	P2	P3	Result
1	7	2	1	+
2	3	5	2	-
3	2	6	2	+
4	6	3	1	+
5	4	4	2	-
6	7	2	1	-
7	2	3	4	+
8	4	3	3	(-)

5. Experiment Results

5.1 Naive Bayes Algorithm Results

Naive Bayes Results

Naive Bayes	Constraint
Accuracy	85%
Time (milli secs)	288
Correctly Classified	85%
InCorrectly Classified	15%

5.2 KNN Algorithm Results

KNN Results

KNN	Constraint
Accuracy	94.1176470588235%
Time (milli secs)	214
Correctly Classified	94.1176470588235%
InCorrectly Classified	5.882352941117648%

6. Conclusion

The placement department plays an important role in student placements, which raises the institute's worth. The current system follows the standard procedure of a firm visiting institutions and doing campus selection. Following the interview procedure, we shall learn the names of the students who have been chosen. However, we require 100% placements in order to improve the college's reputation. As a result, we require a system that can anticipate student placements in advance. As a result, "Student Placement Prediction" application has been created. The system assists colleges in predicting student placement status and boosting placement opportunities.

Future Enhancements

We can employ a larger number of algorithms and apply them to training datasets, allowing us to find the best algorithm. For placement prediction, a greater number of parameters and training datasets can be used.

7. References

[1]. Kohavi, R. and F. Provost(1998) Glossary of terms. Machine Learning 30:271-274.

[2]. Bishop, C.M. (2006) Pattern Recognition and Machine Learning. Springer. ISBN0-387-31073-8.

[3]. Rokach, L and O. Maimon (2008)Data mining with decision trees: theory and applications.World Scientific Pub Co Inc. ISBN 978-98127717711

[4]. Pedregosa, F , G. Varoquax, A. Gramfort, V. Michel, B. Thrion, O. Grisel, M.Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos and D. Cournapeau(2011)Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12: 2825-2830

[5]. Abijith Sankar, P. Divya Bharathi, M. Midhun, K. Vijay, and T. Senthil Kumar, "A Conjectural Study on Machine Learning Algorithms," In Advances in Intelligent Systems and Computing published – Springer - from the proceedings of International Conference ICSCS 2015 2016, Vol.397, pp. 105-116.

[6]. Pal, A.K. and S. Pal (2013)Analysis and Mining of Educational Data for Predicting the Performance of Students. (IJECCE) International Journal of ElectronicsCommunication and Computer Engineering, Vol. 4, Issue 5, pp. 1560-1565, ISSN: 2278-4209, 2013.