INTERNATIONAL RESEARCH JOURNAL OF HUMANITIES AND INTERDISCIPLINARY STUDIES

(Peer-reviewed, Refereed, Indexed & Open Access Journal)

Comparing Accuracy of predictive models for measuring performance of women's in Distance Education - Case Study of HEI'S in India

Mr. Sunil S. Patil

Asst. Professor,
KIT'S IMER, Kolhapur,
(Maharashtra, India)
E-mail: sunilpatil@kitimer.in

Dr. Deelip B. Desai

Asst. Professor,
KIT'S IMER, Kolhapur,
(Maharashtra, India)
E-mail: dbdesai@kitimer.in

Abstract:

Women's education is basic need for a nation to grow in all aspects .The hurdle for women education in India was the man dominating society which was approached from the epics, she have to take care of house and children activities while man have to a take care of economic activities. Distance education has rolled out this barrier. Data was retrieved from kaggle which consists of survey of women's distance education done by MHRD govt. of India in HEI. It consist of 5557 records after cleaning the data. Various predicting algorithm were used for data analysis. The result showed that there is an increase rate of passing percentage and scoring marks in women taking distance education compared to overall percentage of passing and scoring marks. The comparison between models shows that SVM and Random Forest has 100% accuracy.

Keywords: Women's Education, Distance Education, Data Mining, HEI, predictive algorithms.

A. Introduction:

If a woman is educated her whole family gets educated .Women's education is critical for the entire development of the country. A well-educated woman provides the skills, knowledge, and self-assurance required for a family. A well-educated woman is also an asset to a organization she is well-paid at work. Thus there is more   return on investment in education for women than for males. The back-bone for Women's empowerment is education and thus brings a positive attitude change. Educating women gives Basic Right Increases in Literacy Rate and bring Equality in Society for women.

Distance education is only option for this. The benefit of distance education is that it gives women to learn from home. The frequency of attending classes is around 10% of regular mode. This paper analyzes the Performance of women's in Distance Education using various predictive Algorithms.

B. Women education:

As women play a prominent role in the cultural, social, political and economic life of a country and for this Education is the key factor. They are often excluded from education, better, jobs, participation in political system and better health care, decision making, etc.

In India the female literacy rate is 70.3 percent, while the male literacy rate is to be 84.7 percent according to recent survey in lieu with India's average literacy rate 77.7 percent. There is need for constant effort to improve women's education for the vital and overall growth of the country. The deficiency in female literacy is because it was very difficult to get formal education for women's in India due to various hurdles. Women started getting their education in India immediately after the independence but it still was difficult. The solution for this is getting education through distance mode..

C. Distance Education:

Kaplan, Andreas et.al(2016) defines Distance education, also known as distance learning, is the education of students who may not always be physically present at a school or where the learner and the teacher are separated in both time and distance (Anderson, Terry et.al(2020))

Distance education is a type of education that can be effectively incorporated in educating students who are not physically present on site at the classroom. Teachers and students communicate virtually by exchanging printed or through technology.

In recent trends Distance Education is getting more dependent on the online technologies and there is a switching from traditional learning environments or media to online learning environments or media. Thus the term open and distance learning (ODL) has appeared in recent literature especially in the last two years (Aras Bozkurt et.al-2015)

The factors which grow distance education in India are Rapid growth of education, Population, Geographical limitations, Qualification improvement, Desire to work and study, Education for all ages, Financial Circumstances.

D. Women and Distance Education:

Distance education has an important role to play in providing opportunities for women to participate in higher education By adopting Distance education women can equip themselves ready to compete to this burgeoning competitive world, to grow intellectually and   acquire knowledge. Distance Education has been emerged as a nectar to women of all ages.

Some of the barriers in women's learning are Psychological barriers includes Limited career aspirations Institutional Barriers includes bias in recruitment and promotion. Situational Barriers includes Process of Socialization (Patriarchy), Cultural barriers, other barriers include includes curricula that do not fit to the needs of the job, unfriendly learning environment,

The presence of women in distance education is significant not only for the success of women as functionaries in the any of field but also in the growth of the nation.

E. Data mining predictive algorithms:

Some of the most common algorithms in predictive analytics are

- Linear regression.
- Logistic regression.
- Decision tree.
- SVM algorithm.
- Naive Bayes algorithm.
- KNN algorithm.
- K-means.
- Random forest algorithm.

F. Data analysis:

Data analysis was done through two phases

1. Data Cleaning.
2. Data Analysis using Data mining Algorithms.

1. Data Cleaning:

Ministry of Human Resource Development (MHRD), Govt. of India has initiated an All India Survey on Higher Education (AISHE) to build a robust database and to assess the correct picture of higher Education in the country. This dataset contains unit level data from AISHE from 2011-12 to 2015-16.

The initial examination result data set consists of id 14 fields in csv format .It consists of around 2,10,000 records. The 15 fields id   course_mode_id   course_level_id programme_id discipline   appeared_total   appeared_female   passed_total passed_female   broad_discipline_group_id   first_class_passed_total first_class_passed_female   course_id   examination_result_id

It consists of discipline like Science, Geography, Pharmacy, Management, Engineering, Arts, Commerce, Law, etc. The only discipline selected was computers with various sub-disciplines like Computers, M.sc computers, Computer Application and other disciplines irrelevant to study were dropped. Performance of women's in distance education is calculated through result obtained by

female in distance mode examination.

After cleaning of data 5557 records were selected .The irrelevant 4 fields were omitted and 10 fields were selected for further processing of data.

The 10 fields were

id   course_mode course_level_id   discipline   appeared_total   appeared_female passed_total   passed_female first_class_passed_total   first_class_passed_female.

The other 4 fields viz. Female Passing Percent,Total Percent passing, Female First Class Percent, Total Appeared First class Percent were added for analysis purpose.

The comparison chart obtained was

2. Data Analysis using Data mining Algorithms:

The various data mining algorithms SVM, Naive Bayes, k-nearest neighbor, Random forest were applied to cleaned data .The tool used for analyzing the data and applying data mining was Weka 3.7. The results are shown below.

- **SVM algorithm.**

```
=== Summary ===

Correctly Classified Instances         5556              100      %
Incorrectly Classified Instances         0                0      %
Kappa statistic                          1
Mean absolute error                      0
Root mean squared error                  0
Relative absolute error                  0       %
Root relative squared error              0       %
Total Number of Instances             5556

=== Detailed Accuracy By Class ===

                TP Rate   FP Rate   Precision   Recall   F-Measure   MCC    ROC Area  PRC Area  Class
                1.000     0.000     1.000       1.000    1.000       1.000  1.000     1.000     Regular
                1.000     0.000     1.000       1.000    1.000       1.000  1.000     1.000     Distance
Weighted Avg.   1.000     0.000     1.000       1.000    1.000       1.000  1.000     1.000

=== Confusion Matrix ===

    a    b   <-- classified as
 5473    0 |   a = Regular
    0   83 |   b = Distance
```

- **Naive Bayes algorithm.**

```
=== Summary ===

Correctly Classified Instances         5477               98.5781 %
Incorrectly Classified Instances         79                1.4219 %
Kappa statistic                          0.1106
Mean absolute error                      0.0143
Root mean squared error                  0.1184
Relative absolute error                 48.4323 %
Root relative squared error             97.6344 %
Total Number of Instances              5556

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
               1.000    0.940    0.986      1.000   0.993      0.222  0.849     0.997     Regular
               0.060    0.000    0.833      0.060   0.112      0.222  0.849     0.156     Distance
Weighted Avg.  0.986    0.926    0.984      0.986   0.980      0.222  0.849     0.985

=== Confusion Matrix ===

    a    b   <-- classified as
 5472    1 |   a = Regular
   78    5 |   b = Distance
```

- **KNN algorithm.**

k-nearest neighbor algorithm (kNN) on the dataset. In Weka this algorithm is called IBk (Instance Based Learner).

```
Correctly Classified Instances         5542               99.748  %
Incorrectly Classified Instances         14                0.252  %
Kappa statistic                          0.9201
Mean absolute error                      0.0027
Root mean squared error                  0.0502
Relative absolute error                  9.1174 %
Root relative squared error             41.3731 %
Total Number of Instances              5556

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
               0.998    0.012    1.000      0.998   0.999      0.922  0.993     1.000     Regular
               0.988    0.002    0.863      0.988   0.921      0.922  0.993     0.853     Distance
Weighted Avg.  0.997    0.012    0.998      0.997   0.998      0.922  0.993     0.998

=== Confusion Matrix ===

    a    b   <-- classified as
 5460   13 |   a = Regular
    1   82 |   b = Distance
```

- **Random forest algorithm.**

```
Correctly Classified Instances         5556              100      %
Incorrectly Classified Instances          0                0      %
Kappa statistic                           1
Mean absolute error                     0.0089
Root mean squared error                 0.0435
Relative absolute error                29.9756 %
Root relative squared error            35.8886 %
Total Number of Instances              5556

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC    ROC Area  PRC Area  Class
               1.000    0.000    1.000      1.000    1.000      1.000  1.000     1.000     Regular
               1.000    0.000    1.000      1.000    1.000      1.000  1.000     1.000     Distance
Weighted Avg.  1.000    0.000    1.000      1.000    1.000      1.000  1.000     1.000

=== Confusion Matrix ===

    a    b    <-- classified as
 5473    0 |   a = Regular
    0   83 |   b = Distance
```

Sr.no	Algorithm	Correctly Classified Instances	Incorrectly Classified Instances	Accuracy
1	SVM algorithm (Classifier)	5560	0	100%
2	Naive Bayes algorithm(Classifier)	5477	79	98.57%
3	k-nearest neighbor algorithm(Classifier)	5542	14	99.74%
4	Random forest algorithm(Classifier)	5556	0	100%

G. Conclusion:

After analysis of data it was revealed that women taking distance education has high performance than overall performance. It was also revealed that SVM and Random Forest algorithms have high rate of prediction then Naïve Bayes and k-nearest neighbor algorithm as far as this research data is concerned.

References:

1. Bapan Sing, Koushik Acharjee, Subhajit Jana," Development and Obstacles of Women Education in Independent India", Journal of Emerging Technologies and Innovative Research

(JETIR) May 2020, Volume 7, Issue 5

2. Kaplan, Andreas M.; Haenlein, Michael (2016). "Higher education and the digital revolution: About MOOCs, SPOCs, social media, and the Cookie Monster". Business Horizons. 59 (4): 441–50. doi:10.1016/j.bushor.2016.03.008.

3. Honeyman, M; Miller, G (December 1993). "Agriculture distance education: A valid alternative for higher education?" (PDF). Proceedings of the 20th Annual National Agricultural Education Research Meeting: 67–73. Archived (PDF) from the original on 30 April 2015.

4. Anderson, Terry; Rivera Vargas, Pablo (June 2020). "A Critical look at Educational Technology from a Distance Education Perspective". Digital Education Review (37): 208–229. doi:10.1344/der.2020.37.208-229. hdl: 2445/172738. ISSN 2013-9144. S2CID 225664918.

5. Aras BozkurtAras BozkurtEla Akgün ÖzbekEla Akgün ÖzbekSibel YılmazelSibel Yılmazel, Cengiz Hakan AydinCengiz Hakan Aydin "Trends in Distance Education Research: A Content Analysis of Journals" January 2015International Review of Research in Open and Distance Learning 16(1):330-363, DOI: 10.19173/irrodl.v16i1.1953

6. Bapan Sing, Koushik Acharjee, Subhajit Jana, "Development and Obstacles of Women Education

in Independent India", Journal of Emerging Technologies and Innovative Research (JETIR), May 2020, Volume 7, Issue 5.

7.Afa-Dul Mujiaba, "Empowerment of women through distance education in India" , Research Journal of Humanities and Social Sciences, ISSN 0975-6795,2011Vol- 2, Issue 4.

8. Data Set Reference- https://www.kaggle.com/.