

Analysis of Classification Techniques for Prediction of Tuberculosis Defaulters

Brian Mc George
University of Cape Town
Cape Town, South Africa
mcgbri004@myuct.ac.za

ABSTRACT

1. INTRODUCTION

In 2013 over 210 000 patients defaulted from Tuberculosis (TB) treatment. South Africa has the second highest incidence rate of Tuberculosis (TB) per 100 000 of the population [4]. Its neighbour, Lesotho, has the highest rate of TB per 100 000 of the population [4]. South Africa has the sixth highest incidence rate of TB in terms of absolute occurrences [4]. The consequences of defaulting TB treatment include: increased drug resistance, increased health system costs [2, 3], higher risk of mortality, continued risk of transmitting the disease to others [2] and increased rate of recurrent disease [1]. The spread of TB can be reduced if the individuals who have a high risk of defaulting can be predicted. This will also reduce health system costs.

1. Discuss datasets used and possibly expansion of TB issues in those specific counties (Peru and Malawi)
2. Outline briefly how the datasets are used and that each technique is also benchmarked against the well known Australian and German financial to determine how applicable credit scoring research is to TB default prediction for the two TB datasets.
3. Link TB classification to credit scoring and outline notable similarities and differences
4. Summarise overall paper

2. RELATED WORK

1. Summarised version of original literature review except the parts using temporal aspects
2. Summarise what has been done determining TB predictors (which is a different aim)
3. Summarise what has been done in the credit scoring financial space as a lot of research has been done in this field.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

3. METHOD

1. Outline testing procedure and optimisation strategy for each classification technique. To optimise each classification technique, a grid search is conducted across a reasonable parameter space. For each parameter set, the results are averaged over 3 runs, this is done because of the stochastic nature of initialising the training of each classifier.
2. Brief overview of each classification technique and possibly give a reason why this technique may work well for our application

4. RESULTS

1. Tables with true positive, true negative, false positive and false negative rate for each classifier on each dataset
2. ROC curves which can be used to determine true positive for an acceptable amount of false positives

5. DISCUSSION

1. Discussion on best classification technique
2. Outline similarities and differences in results between the two TB datasets and between TB and financial datasets and determine if they are similar enough that results for the German and Australian credit scoring datasets could be used for the two TB datasets.

6. FUTURE WORK

1. Likely something along the lines of utilising more temporal based data so that classification is not just done at registration but also at each check-up for example. Future work may also be the testing of more classification techniques as well as datasets from other parts of the world.

7. CONCLUSIONS

8. REFERENCES

- [1] U. M. Jha, S. Satyanarayana, P. K. Dewan, S. Chadha, F. Wares, S. Sahu, D. Gupta, and L. S. Chauhan. Risk factors for treatment default among re-treatment tuberculosis patients in india, 2006. *PLoS ONE*, 5(1):1–7, January 2010.
- [2] B. Lackey, C. Seas, P. Van der Stuyft, and L. Otero. Patient characteristics associated with tuberculosis treatment default: A cohort study in a high-incidence area of lima, peru. *PLoS ONE*, 10(6):1–11, 2015.
- [3] B. Muture, M. N. Keraka, P. K. Kimuu, E. W. Kabiru, V. O. Ombeka, and F. Oguya. Factors associated with default from treatment among tuberculosis patients in nairobi province, kenya: A case control study. *BMC Public Health*, 11(1):696–105, September 2011.
- [4] World Health Organisation. Global tuberculosis report 2015.