CHAPTER 8

CONCLUSION AND SCOPE FOR FUTURE WORK

Data mining is the concept of extracting knowledge from large dataset. The proposed method uses classification techniques to classify the large dataset in short time with more accuracy. This method comes under the supervised learning system where the class is already known and the new method produced accurate result even for a huge dataset. Since the dataset is growing every day, it becomes necessary for any mining process to extract the results as early as possible. The proposed method used share market dataset for classification to predict those shares which would give more profit in near future. The proposed method helps the investor to choose a profitable share to invest and earn profit.

The objective of this research is to classify the shares for investor to gain profit while they are going to invest in share market using the FFC. The performance of FFC is evaluated with the following parameters:

- Classification time
- Accuracy
- Sensitivity
- Specificity

The proposed methodology developed an efficient fuzzy fast classification method that accurately classifies the large dataset in a smaller time compared to other methods. The research is carried out in five phases.
- Phase 1: Preprocessing database.
- Phase 2: Attribute reduction using entropy.
- Phase 3: Classification of dataset with MSVM.
- Phase 4: Extraction of complex records from boundary analysis.
- Phase 5: Fuzzy Fast Classification (FFC) for error records.

The raw dataset is preprocessed with various methods to make a model suitable for classification process. A standard step to data preprocessing is normalization. The standardization feature used for the share market dataset is rescaling since the values may be in lakhs and millions. This makes it necessary to perform non-zero analysis as rescaling might have zero values that leads to misclassification and the result of accuracy might go down. Next the attribute reduction is carried out as the data could contain many redundant or irrelevant features and this reduces the effectiveness of data mining. Also the attribute selection is more important for any data analysis which is done with the Shannon’s entropy in the proposed method.

The proposed method MSVM has been used to classify the entire dataset with mean separation. The dataset obtained after the preprocessing and attribute selection are calculated with weight for each record then the records are sorted based on the calculated weight and the mean value is calculated. The values below the mean is assigned as class1 (low) and the values above the mean is assigned as class2 (high). Here the classification may be faster, but there appears to be more error near the center of database. Hence, the boundary region is fixed and complex records are extracted for further classification. The upper and lower boundary region are fixed based on the center of the dataset with the condition that difference between class1 and class2 are calculated and depending upon the percentage of difference
with overall dataset, the complex regions were fixed and boundary regions are extracted.

The proposed method FFC considered only the error records from the complex record and leaving out the records that are classified correctly. These error records are classified using the fuzzy classification by calculating the degree of membership value for each attribute. Once the degree of fuzzy values are calculated using the upper and lower bound for each attribute, then either class 1 low or class 2 high is assigned for each attribute and finally fuzzy logic is applied by combining the attributes. The results show that the proposed method had improved accuracy and classification time.

This method FFC helps the stock investor to choose an appropriate share market at any point of time. Since the stock market is always changing over time and a right decision cannot be made at a single sight. So the proposed approach with previous information classifies the dataset at a faster rate that helps the investor to identify right shares at right time so as to gain more profit by investing in the appropriate share. The experimental results show that FFC approach has an improved performance when compared to other algorithms. This approach could be applied to any huge dataset for performing classification at a faster rate.

**PERFORMANCE EVALUATION**

Experiments are conducted in Pentium IV machine with 4GB RAM. The proposed method shows an improved classification performance compared to other methods. The performances of the proposed approaches are evaluated using the following parameters:

- Classification Time
- Accuracy
Sensitivity

Specificity

With all these parameters the proposed FFC had better performance than other methods like SVM, LDA and LIBSVM.

SCOPE FOR FUTURE ENHANCEMENTS

A future enhancement could still increase the sensitivity, specificity and accuracy of algorithm by correcting the errors outside the complex region. This work deals with binary classification considering the investor could only make a decision whether to invest in the share market or not. This could be further enhanced to three or more choices in class as very low and very high so that the investor could make better decision.

The proposed work considered only the numerical dataset. Some dataset may contain images like graph representation of share market which are not considered in the method. Hence a future enhancement to this research work could be image or graph type datasets that might be challenging as it could take longer time for classification.