CHAPTER 7

CONCLUSION

The proposed FPED technique demonstrates that it provides accurate data records by eliminating the errors such as, duplicate records, near duplicate records, misspelling errors and illegal value errors. After completing some of preprocessing steps, Employed Rabin’s 32-bit fingerprinting algorithm for removing duplicate records which successfully removed the duplicate records, Levenshtein distance is successfully cleansed the dataset from near-duplicate records and then subsequently detect the illegal errors and rectified it with the help of rule base method. For experimental results, the synthetic data was taken from a directory containing contact number details of clients and the evaluation results have showed that FPED technique have very high recall and very low Reverse False-Positive error (RFP) and False-Positive Error (FPE) and thus, obtained a dirt free and consistent dataset.

The obtained result can also be optimized using optimization techniques such as, selecting the best set of features from the initial sets and fuzzy classifier may employed to detect of redundant data. This research work can be make utilized for the effective decision making, management and processing such as unique identification card, budget allocation made by government based on census, voter identification card for polling system, issuing ration cards, medical insurance schemes, other schemes which may have chances of duplication and illegitimate errors. Also, this cleaned data can be used for predicting or confirming the disease, doing medical diagnosis for the rural areas which may have lesser equipment facilities such as CT, MRI scans and other lab amenities.