

## **55. STORAGE OPTIMIZATION IN CLOUD BACKUP SERVICES USING RECORD LINKAGE**

M.LATHA, P.NEELAVENI

bharathilatha91@gmail.com, srirang.neels@gmail.com

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

G.K.M. College of engineering and technology

Storage optimization is an essential task in many domains. Record linkage is identifying different data items that refer to the same entity across different data sources. The goal of the record linkage task is joining data sets that do not share a common identifier. One-to-many and many-to-many record linkage is an essential task in many domains, yet only a handful of prior publications have addressed this issue. Furthermore, while traditionally data linkage is performed among entities of the same type, it is extremely necessary to develop linkage techniques that link between matching entities of different types as well. Record linkage scenarios include: linking data when combining two different databases and data deduplication. We present one-class clustering tree (OCCT) that characterizes the entities that should be linked together. The tree is built such that it is easy to understand and transform into association rules, i.e., the inner nodes consist only of features describe the first set of entities, while the leaves of the tree represent features of their matching entities from the second data set. Four splitting criteria and two different pruning methods which can be used for inducing the OCCT. The method was evaluated using data sets from three different domains. In addition, while record linkage is usually performed among the same and different types of entities. Maximum likelihood estimation can be used to determine the probability of record pair match and also learning the characteristics of the matched records. After that, the matched record pair linked using one class clustering tree.

Index Terms- Cloud Computing, Clustering, Classification, Storage optimization, Decision tree induction.

*Journal of Science and Innovative Engineering & Technology*