International Journal For Technological Research In Engineering
Volume 5, Issue 9, May-2018 *ISSN (Online): 2347 - 4718*

TAXONOMY CLASSIFICATION OF PRODUCT USING MACHINE LEARNING TECHNIQUES

Shylashree U R[1], Kiran B N[2]
[1]Student of Computer Network Engineering, [2]Assistant Professor
Department of ISE, The National Institute of Engineering, Mysuru-570008

ABSTRACT: Product Description must be classified. Each product will be having its own taxonomy, which organises data into its category. Manually classifying these products is not a trivial task, it require lot of domain expertise and knowledge on product domain. This paper describes the method to classify the products automatically to its respective categories using Machine Learning Technique. That is given product description that includes Brand name, invoice description, catalog description etc.. Its category will be predicted. At the same time paper also describe detail analysis of method and few challenges that will come across during implementation. Our implementation results show consequential progress over Standard results. Taking into particular criteria, our implementation is potentially able to considerably increase automation of categorization of products.

Keywords: Machine learning, Content Management, Taxonomy Classification, E-commerce

I. INTRODUCTION

Machine learning is a strategy for information investigation that robotizes diagnostic model building. Machine learning methods work for automatic categorization of products. Product categorization for E-business destinations is a mainstay for successful marketing and offer of products recorded on several online stores like Amazon, eBay etc. Product categorization is the assignment of mechanically predicting a taxonomy lane for an item in a predefined taxonomy specified in a written item depiction. The model would have the capability to classify the product's category and their subcategory based upon the product data. This is valuable for the situation where a business has a rundown of new products that they need to automatically classify the products. Product categorization is an unambiguous classification of products in various product classes. Each class has a few equivalent words, in this way finding the correct product is troublesome for everybody. Supervised learning methods are relevant in various spaces. In all aspects of the world, there are distinctive frameworks of product categorization. The idea of product categorization comprises of separating products as indicated by particular attributes with the goal that they frame an organized portfolio. When all is said in done, makers utilize an informal product categorization framework yet there are likewise many standardized strategies for product categorization manufactured by different industry associations.

Product categorization, the assignment of categorizing arriving product offers from online sites as per pre-characterized item taxonomy. Here use supervised learning technique as an approach to automatically classify products. It makes the way toward discovery what you are searching for at ease. There are a few difficulties that should be overcome so as to build a robust product classifier, beginning with building up a consistent taxonomy and a data set sufficient for validation. The exploratory outcomes demonstrate that the proposed framework fills in as an effective product classifier and also this paper explains about automatic product classification which includes steps such as pre-processing, vector representation and efficient algorithm is used for prediction of the class. The development of the tool is mainly for classification because every month nearly 1 lakh product gets added to the database. To classify products manually it needs experts and approximately 5000 items can be classified in a day, it is more time-consuming. The main aim of the classifier is to classify the products to their respective class with reasonable time so classifier helps to overcome the issues.

II. PROPOSED SYSTEM

Little to medium measured organizations who offer items online spend a critical piece of their opportunity, cash, and exertion sorting out the items they pitch, understanding shopper conduct to better market their items and figuring out which items to offer that is, given a product with accompanying informational details, product should automatically classify into a particular category with similar products, e.g., 'Hardware' or 'Car'. You can use R language to do that work.

Architecture of the Product categorization

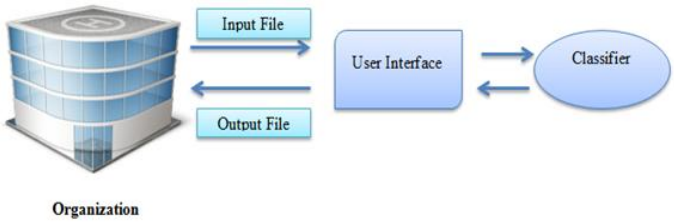

Figure 1: System Outline

Figure 1 describes about outline of system that consists organization. The organization essentially offers support and assets that quicken development and to remain in front of the opposition. The company has the dataset of the distributors, manufacturers and industrial products this has to be categorizing with respect to their classes. User interface selects the input file where the taxonomy classification operation takes place in the classifier, giving the categorized products in the output file.

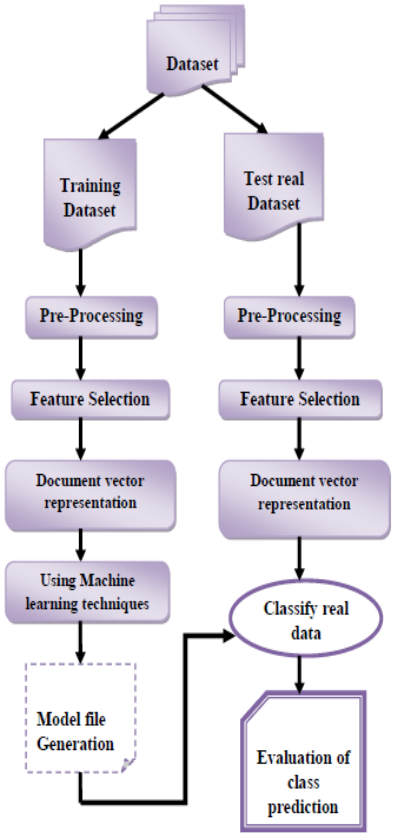

Figure 2: System high level Architecture.

Machine learning iteratively analyses the data using algorithms this automates to build the model. This enables computers to discover shrouded familiarities without being clearly programmed where to look and make a prediction on data. The problem is to determine the taxonomy of the products where there is more number of classes hence it is one of the difficult tasks to classify. Using supervised learning procedures studies the training data and generates a model file which can be utilized for mapping new illustrations. It is a two-step process, the first step is model file generation next is using model file prediction of new data this is explained in the figure 2.

III. III IMPLEMENTATION

Pre-processing: Pre-processing or Noise removal technique is a critical step in machine learning. It will enhance the quality of data. Due to the difference in representation of information, not all data in the product description are required to classify the products. Pre-processing techniques removes the feature or words which are not playing a vital role in classification. It includes stop words removal, removal of punctuation, removal of numbers, and removal of reserved words, removal of special character and also stemming, lemmatization, stripe whitespace. Reserved words (while, for, if) are the words that can be used as an identifier, few machine learning technique cannot handle those words as a feature.

Feature Selection: Features are basically words used in describing the products which play a critical role in the classification. When large set of products is considered, it results in high dimensional feature space which causes both space and time complexity. An enormous percentage of the features are not significant along with valuable for categorizing the products also noise features might pointedly lessen the precision. Attribute selection is the progression of picking a subsection of features which are having the highest score. One more profit of feature selection is its propensity in the direction of diminish curve over fitting. Scores for features are given by using feature evaluation metrics like information gain, Chi-square and gain ratio, etc.

Vector Representation: After features selection, construct the vector space representation which is basically represents the frequency of each of the terms in the given document wherever every feature happens in any event once in a specific least number of reports, which improves the expansible of a product classifier.

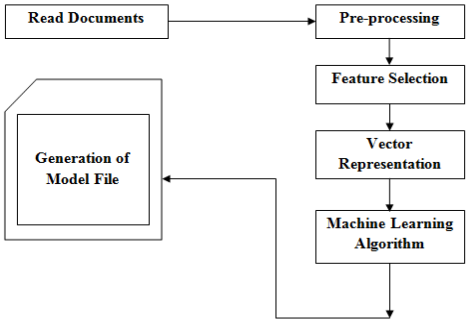

Figure 3:Flow Chart of Model File Generation

Algorithm selection: The random forests algorithm is solitary of the finest among categorization algorithm in a supervised learning, ready to arrange a lot of information with exactness. Random forests are a blend of tree indicators where every tree relies on upon the estimations of an arbitrary vector inspected freely through a similar dispersion of every tree in the forest. The essential rule is, a gathering of feeble learners be able to meet up toward frame a solid learner. These are a great device for constructing expectations taking into account they don't over fit on account of the rule of huge information. Presenting the correct sort of haphazardness make them exact classifiers. After prediction note down the classified with the predicted class name, product description and other information to excel file.

IV. CONCLUSION AND FUTURE WORK

In this work we proposed a classification tool points on motorizing the automatic categorization progression of manufactured goods information. Precision rates in the vicinity of 80% and 90% demonstrate to this procedure be capable of motorized to a quantity where extreme price lessening be able to accomplish which is a pre imperative for adaptable e-business. The achievement of tool depends on top of the blend of pre-processing, vector representation and machine learning techniques. Item index, item standard, and item depictions are portrayed in different dialects are an extraordinary need for E-business. As of now, the device bolsters English. A development towards extra lingo is a

basic utilized for open and in addition versatile e-business and also Multi standard order is a basic concern utilized for open and additionally versatile e-business. At introduce, commercial centres develop single example to build data of their customers reachable

REFERENCES

[1] Sushant Shankar and Irving Lin "Applying Machine Learning to Product Categorization", Department of Computer Science, Stanford University

[2] Aurangzeb Khan, Baharum Baharudin, Lam Hong Lee, Khairullah Khan,"A Review of Machine Learning Algorithms for Text Documents Classification", Journal Of Advances In Information Technology, Vol. 1, No. 1, February 2010

[3] Y. Ding, M. Korotkiy, B. Omelayenko, V. Kartseva, V. Zykov, M. Klein, E. Schulten, and D. Fensel," GoldenBullet: Automated Classification of Product Data in Ecommerce",In Proceedings of Business Information Systems Conference (BIS 2002)

[4] M. Ikonomakis, S. Kotsiantis, V. Tampakas, "Text Classification Using Machine Learning Techniques", WSEAS TRANSACTIONS on COMPUTERS, Issue 8, Volume 4, August 2005, pp. 966-974

[5] Chong Sun, Narasimhan Rampalli, Frank Yang , AnHai Doan," Chimera: LargeScale Classification using Machine Learning, Rules, and Crowdsourcing", Sun et al. 2014 (WalmartLabs)

[6] Yang Song, Ming Zhao, Jay Yagnik, and Xiaoyun Wu," Taxonomic Classification for Web-based Videos", Google Inc., Mountain View, CA 94043, USA.

