

IMPROVING ACCURACY OF CLASSIFYING UNSTRUCTURED DATA WITH MACHINE LEARNING

Customer Story | Online Aggregation Services



CLIENT

The client is a global procurement facilitator for tenders. The company provides information on local & international government tenders from all major sectors like construction, defence, transportation, medical & healthcare, energy, power & telecommunication, etc. Information on government tenders, tenders by organization and invitations to receive tenders from all over the world, is provided online. Additionally, the company provides government projects and contract awards details from all sectors globally.

CHALLENGE

SEMI-AUTOMATIC
CATEGORIZATION OF
UNSTRUCTURED DATA WITH
LOW ACCURACY

HEAVY DEPENDENCE
ON MANUAL LABOUR

The client received an average of 60,000 tenders a day. Each tender could fall into any of 61 main categories and 8300 sub categories. Moreover, there could be multiple categories or products bundled in a single tender. The problem was further complicated with the fact that the tender data was extremely noisy with sudden mentions of words from other languages like Hindi, Dutch, Swiss etc. The data was unstructured and meaning had to be derived from long and detailed descriptions of the tenders.

The client's tender classification system was a semi-automatic one and used keyword matches in legacy systems. This system didn't seem to work or scale well, and required additional hours of manual labour. They had a very limited corpus (vocabulary) of words compared to the the keywords that appeared in the tenders. The solution required building a larger and all-encompassing vocabulary for them as well.

SOLUTION

**AUTOMATED CLASSIFICATION
USING MACHINE LEARNING**

**DEEP LEARNING
USING VECTORIZATION**

Using machine learning, our algorithm learned to classify these tenders into the correct main category first. After predicting the main category, it would look for the correct sub-categories within that main category. The algorithm used a combination of keyword search and contextual learning of words to increase accuracy. We used a combination of traditional approaches combined with cutting-edge deep learning using vectorization to get a better output.

BENEFITS

**90% REDUCTION
IN MANUAL LABOUR**

**ACCURACY OF 87% WITH
ALGORITHMIC CLASSIFICATION**

With our machine learning based solution, the manual labour involved was reduced by 90%. The algorithm scaled well to new unseen tenders because of its word-based context capturing ability. Our algorithm achieved an accuracy of about 87% after the first training. The client now uses this algorithm and the misclassified tenders are used as an input for re-training the model, thereby further improving accuracy with every re-training process.

ABOUT QUOSPHERE

Quosphere provides innovative solutions in the areas of Business Intelligence, Big Data, Data Visualisation, Analytical Frameworks and Cloud Technology. Our services are focused on helping organisations translate the effectiveness of Big Data & Machine Learning technologies into great business impact.

Quosphere has supported businesses across the world in increasing productivity through customized services and world-class business-cycle solutions. Setting standards in data analytics across spectrums with advanced technology and dedicated workforce has been our key achievement.

GET IN TOUCH WITH US

MUMBAI

1004, 10th floor, Rupa Solitaire,
Millenium Business Park, Navi Mumbai,
Maharashtra 400710, India
Phone: +91-22-41031111

NEW YORK

19 W 34th St #1018,
New York, NY 10001
Phone/Fax: 001-646-394-9093