# Mining Competitors from Large Unstructured Datasets

Abstract:

Along line of research has demonstrated the strategic importance of identifying and monitoring a firm's competitors. Motivated by this problem, the marketing and management community have focused on empirical methods for competitor identification as well as on methods for analyzing known competitors. Extant research on the former has focused on mining comparative expressions (e.g. Item A is better than Item ) from the Web or other textual sources. Even though such expressions can indeed be indicators of competitiveness, they are absent in many domains.

**Existing System:**

To the best of our knowledge, our work is the first to address the evaluation of competitiveness via the analysis of large unstructured datasets, without the need for direct comparative evidence. Nonetheless, our work has ties to previous work from various domains.

**Managerial Competitor Identification:** The management literature is rich with works that focus on how managers can *manually* identify competitors.

**Finding Competitive Products:** The first step in these approaches is the definition of a dominance function that represents the value of a product.

**Skyline computation:** Our work leverages concepts and techniques from the extensive literature on skyline computation.

Disadvantages:

- Dependency on transactional data.

**Proposed System:**

In this paper, we present a formal definition of the competitiveness between two items, based on the market segments that they can both cover. Our evaluation of competitiveness utilizes customer reviews, an abundant source of information that is available in a wide range of domains. We present efficient methods for evaluating competitiveness in large review datasets and address the natural problem of finding the top-k competitors of a given item. Finally, we

evaluate the quality of our results and the scalability of our approach using multiple datasets from different domains.

Advantages:

- Efficient and applicable to domains with very large populations of items.

**Modules:**

- Finding The Top-K Competitors.
    1. CMiner Algorithm.
- Boosting The CMiner Algorithm.

## SYSTEM REQUIREMENTS

### H/W System Configuration:-

| | | |
|---|---|---|
| Processor | : | Pentium –III |
| RAM | : | 256 MB (min) |
| Hard Disk | : | 20 GB |
| Key Board | : | Standard Windows Keyboard |
| Mouse | : | Two or Three Button Mouse |
| Monitor | : | SVGA |

### S/W System Configuration:-

| | | |
|---|---|---|
| Operating System | : | Windows95/98/2000/XP |
| Application Server | : | Tomcat5.0/6.X |
| Front End | : | HTML, Jsp |
| Scripts | : | JavaScript. |
| Server side Script | : | Java Server Pages. |
| Database | : | MySQL 5.0 |

Database Connectivity : JDBC