



## Cost Based On Product Quality Prediction Using Datamining

S. Veeramalai<sup>1</sup>, Mr. T. Praveen<sup>1</sup>, S. Pradeepa Natarajan<sup>2</sup>  
<sup>1</sup>Assistant Professor, <sup>2</sup>Student

<sup>1</sup>Department of Computer Science Engineering, <sup>2</sup>Master of Computer Application Department  
Veltech Hightech Dr Rangarajan Dr Sakunthala Engineering College  
Affiliated to Anna University, Chennai, Tamil Nadu, India

### ABSTRACT

In real time problem is the finding the top-k profitable product in the existing market. To find in the best product in the market such the new product could not dominate by other product in the market. Authentication and skyline analysis are best decision-making applications. In the existing focus the customers to analysis the set of product from a pool of given products. Set of product in the existing market want to identify possible product such that novel product are not dominated by other in market. There are two problem occasion of the finding top-k preferable products are studied. In first problem instance to set the cost of these products maximized number of total in income.

**Keywords:** prediction engine, top-k product

### 1.Introduction:

#### Skyline:

Dominance and skyline analysis has been well recognized in cost based on product quality prediction engine. A package which is not dominated by any other packages is said to be a skyline package or it is in the skyline. The packages in the skyline are the best possible tradeoffs between the two factors in question.[2]

The skyline operator is important for several applications involving decision making. Skylines are related to several other well-known problems, top-K queries and nearest neighbor search.

### 1.1.Existing System

To find top k profitable products a inexpert way for this illustration problem is to itemize all possible subsets of size k from the obtainable set and then analyze the sum of the profits of each possible subset and finally select the subset with greatest sum.

To find top k popular products is an immature way for this instance problem is similar to that of the first instance.

First find all possible subsets of size k from the available set and then choose the subset with greatest number of customers.[3]

### Existing Algorithm:

Present algorithms to find top-k totally and partially unexplained sequences and classes.

For ease of presentation, we assume  $|f.obs| = 1$  for every OID f in an observation sequence (this makes the algorithms much more concise generalization to the case of multiple action symbols per OID is straightforward.[4]

Given an observation sequence  $v = (f_1; \dots; f_N)$ , we use  $v(i; j)$  ( $1 < i < j < n$ ) to denote the sequence  $S = (f_i; s_i); \dots; (f_j; s_j)$ , where SK is the only element in  $f_k.obs$ ,  $i < k < j$ .

**Disadvantages of Existing System:**

This way used to finding top k profitable products (first problem instance) is not scalable because there are an exponential number of all possible subsets.

These procedures used for finding top k popular products (second problem instance) are also not scalable because this also involves finding all possible subsets which is exponential.

**1.2. Proposed System**

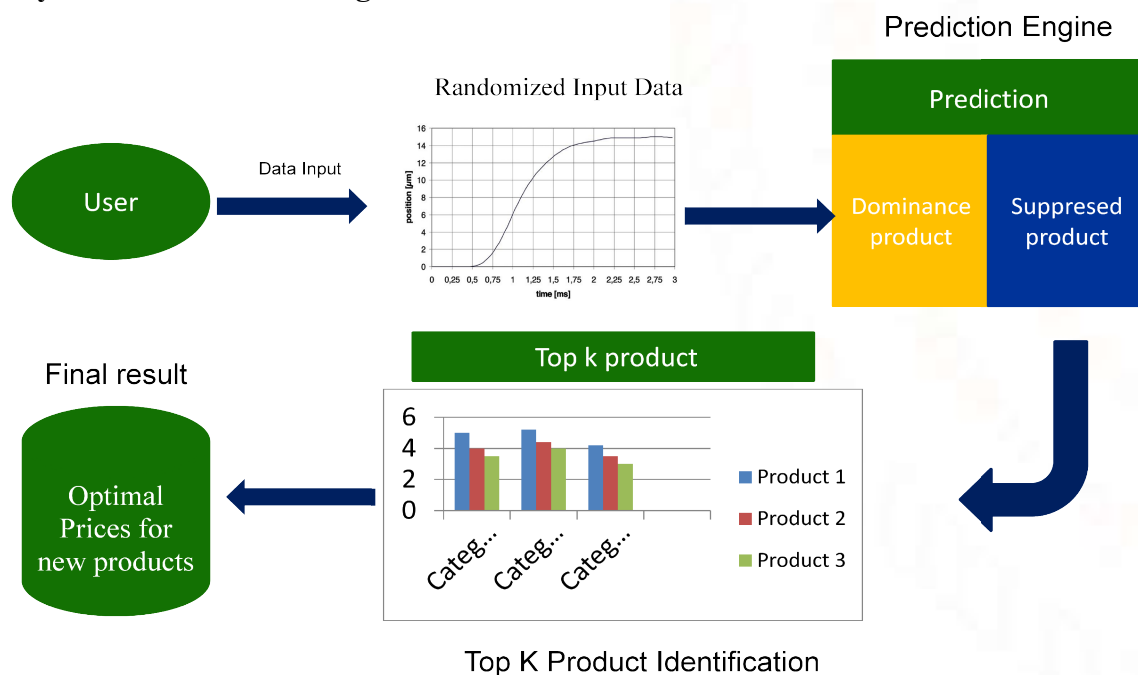
**To find top k profitable products:**

In proposed system the dynamic programming approaches, which finds an optimal solution when there are two attributes to be considered. Here we are utilizing the option of find Optimal Incremental Property Algorithm. In which, we are trying to authenticate/discover the quasi dominance of the products and apart from that our system will recognize the skyline checks on the available data. Based on an optimized check among the two methodological jargons, the profitable products will be identified.

In this module, we are using the attributes as the main criteria to define and decide the Top K Products. In case of preferable products, we have taken the summation of the user ratings as the criteria to identify the best products. Over Profitable products; our system utilizes the car cost, duration of the car and user rating. Based on the mixed summarization and summation, we have identified the Profitable products in the category.

The adaptive pulling strategy related to the products prioritizes access among the two relations based on the observed data. [3]The main idea behind this approach is to read the tuples from a relation only if there is possible evidence regarding the new tuples which will help and satisfy the termination condition. Intuitively, this prioritization process helps the algorithm terminate faster and sooner, thus improving its performance. Obviously, the popular products based on skyline processing are done.

**2. Proposed System Architecture Diagram**



In this architecture the user can input the rating for product, and other product detail also input in the data. Prediction engine gather the information from the user rating and other product detail. And analysis the both data by using prediction, then separate the dominance and suppressed product. Dominance product means high level product in market level and

suppressed product means low level product. In dominance product find the top k product and analysis product quality and cost for the product.

**Advantages of Proposed System**

For the first problem of finding top-k profitable products, a dynamic programming approach which

can find an optimal solution when there are two attributes to be considered is proposed.

An incremental approach is used to handle dynamic datasets that change over time.

### 3. Literature Survey:

#### **TITLE: Mining top-K frequent item sets from data streams**

The difficulty of mining top K frequent item sets in data streams is taken care in this project. The author has introduced a method based on the Chernoff bound with an assurance of the output quality and also a bound on the memory usage.

The author has proposed an algorithm based on the Lossy Counting Algorithm. In most of the experiments of the two proposed algorithms, we obtain perfect solutions [5] and the memory space occupied by our algorithms is very small.

#### **MERITS**

This paper proposes “Chernoff bound” which is a bound on the probability that an arbitrary variable deviates by a certain amount from its expectation.

This algorithm clearly focusses on the memory usage; “unpromising item sets will be pruned regularly to keep the memory usage low” is the basic underlying concept of this paper.

#### **DEMERITS**

The author didn't focus on the performance of the system and accuracy of the data retrieved.

#### **TITLE: Mining Top-K Patterns from Binary Datasets in presence of Noise**

This project directed on the discovery of patterns in binary dataset in many of the applications, e.g. in electronic commerce, TCP/IP networking, Web usage logging, and etc. information on some of the factors like: overlapping vs. non overlapping patterns, presence of noise, and extraction of the most important patterns only.

In this paper the author have formalize the problem of discovering the Top-K patterns from binary datasets in presence of noise, as the minimization of a novel cost function.

According to the minimum Description Length principle, the proposed cost function favors succinct

pattern sets that may approximately describe the input data. [6]

#### **MERITS**

This project determines the exact top level data from the binary datasets even in the presence of issues with the data.

They utilized the PANDA algorithm to identify the level of noise and rectification technique on the noise data.

#### **DEMERITS**

The author didn't clarify the factor of specifying the level of noise identification in this project which is a great drawback of this system.

#### **TITLE: Mining and Representing User Interests the Case of Tagging Practices**

In this paper, we provide a novel approach for clustering user-centric interests by analyzing tagging practices of individual users.

The FCA (Formal concept analysis) and a significance measure are based on the weight of the tags in the given data set.

The concept analysis technique makes it easy to mine common tags with respect to users in the data set.[7]

#### **MERITS:**

The approach can be used to suggest new social relationships within a small-size group based on the users' interests.

It is easy to aggregate user interests from multiple sources.

#### **DEMERITS:**

It is not straightforward to build general-level information for the given data.

We can't carry out the building of large community-level informations by adopting the approaches.

## **4.METHODOLOGY**

### **Find Optimal Incremental Property Algorithm**

- Input: A set  $Q_{i-1} (= \{q_1, q_2, \dots, q_{i-1}\})$ , tuple  $q_i$  in  $Q_0$  and the optimal price assignment vector  $v_{i-1}$  of  $Q_{i-1}$

- Output: the optimal price task vector  $v_i$  of  $Q_i$  with the h-dominance constraint

The relational database could contain too many disconnected components, in which case our link analysis approach is almost useless.

- Deletion of an existing package
- Insertion of a new package
- Modifying the attribute values of an existing package

Method are used to analysis the data in the database which are identify the top-k preferable and profitable product.

- Authentication Module
- Product Details Screen

### To find top k popular products:

#### PROPOSED SYSTEM – EXAMPLE 1

##### Profitable price with one package

- If price is \$100 profit is Null
- If price is \$400 Profit is \$300 but dominated by P2
- If price is \$300 Profit is \$200; Also it is not dominated by any other package

Package	Destination to beach	Price
P1	7.0 units	200
P2	4.0 units	350
P3	1.0 unit	500
P4	3.0 units	600

Package	Destination to beach	Cost	Price
Q1	5.0 units	100	?
Q2	4.5 units	200	?
Q3	0.5 units	400	?

### 6. Conclusion

In this paper, how the user find the Top-k profitable as well as preferable product are not dominated by any other product existing in the market. This work proposes to choose the best product to get maximum profit i.e. multiple decisions are arise when select the product, even though it suggests best product to get maximum profit and also all the problems for finding Top-k profitable product are solved and synthetic data has been used the result obtained are practically and theoretically accurate for testing. It is also suitable for real time data sets (i.e. companies real time trading records connected to server).

### REFERENCES:

1. Progressive skyline computation in database systems JULY 18, 2016
2. Qian Wan<sup>1</sup>, Raymond Chi-Wing Wong<sup>1</sup>, Ihab F. Ilyas<sup>2</sup>, M. Tamer Ozsu<sup>2</sup>, Yu Peng<sup>1</sup> Creating Competitive Products

- Top K Products Module
- Top K Preferable/Profitable Module
- Product Cost Prediction Module

### 5. Results and Discussion:

The External source is (i.e. shares have to be either sold or purchased) added to Database. Data Administrate sends top level product. Now the registers his/her name and then logs into product database. The share portal will display top-k (according to algorithm) profit product along with their profit value in user account as shown in Fig.3 these are our required profitable and popular product. Now the user can decide to fix the optimal prices for productand making better profit in that product.

#### PROPOSED SYSTEM – EXAMPLE 2

##### Profitable price with two packages

- Prices of Q1 and Q2 are \$300 each
- But Q2 dominates Q1
- Optimal price assignment: Price of Q1 = \$250
- Price of Q2 = \$300

• Total profit = \$150 + \$100 = \$250

Package	Destination to beach	Price
P1	7.0 units	200
P2	4.0 units	350
P3	1.0 unit	500
P4	3.0 units	600

Package	Destination to beach	Cost	Price
Q1	5.0 units	100	?
Q2	4.5 units	200	?
Q3	0.5 units	400	?

3. Mr.C.Sivakumar<sup>1</sup>, Mr.M.Arulprakash A Framework Outline for Word Catalog Based Model in Cloud
4. Preethi.P Pradeep.R “Price prediction System using Data Anonymity and Top-k Products” Issue 1 January 2015
5. Raymond Chi-Wing Wong, Ada Wai-Chee Fu Mining top-K frequent item sets from data streams Published 2006 in Data Mining and Knowledge Discovery
6. Claudio Lucchese, Salvatore Orlando, Raffaele Perego Mining Top-K Patterns from Binary Datasets in presence of Noise Published: 2010
7. 1E. Munuswamy, 2M.Ferni ukrit Efficient Secret Data Sharing Using Multimedia Compression Paradigm © 2015 IJEDR | NC3N 2015 | ISSN: 2321-9939

8. Mary Sowjanya “Multi- criteria Decision making for Identifying Top-k profitable Stocks from Stock market” 11 November 2016
9. Dhanaji Jadhav Using Dynamic Constraints Find Top Products April 2015
10. B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, “Item-based collaborative filtering recommendation algorithms,” in Proc. WWW, 2001, pp. 285–295.
11. J. Herlocker, J. Konstan, and J. Riedl, “An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms,” Inform. Retrieval, vol. 5, no. 4, pp. 287–310, 2002.
12. M. Deshpande and G. Karypis, “Item-based top-n recommendation algorithms,” ACM Trans. Inform. Syst., vol. 22, no. 1, pp. 143–177, 2004.
13. R. Bell, Y. Koren, and C. Volinsky, “Modeling relationships at multiple scales to improve accuracy of large recommender systems,” in Proc. KDD, 2007, pp. 95–104.
14. P. Bedi, H. Kaur, and S. Marwaha, “Trust based recommender system for semantic web,” in Proc. IJCAI, 2007, pp. 2677–2682.