

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 3, March 2019

## Skyline Product Combination for Customer Favorable Products in Shopping

Jaya preetha.B<sup>1</sup>, Nivetha.A<sup>2</sup>, Priyanka.V<sup>3</sup>, Vishnu priya.M<sup>4</sup>, Dinesh.G.V<sup>5</sup>

UG Student, Department of Computer Science and Engineering, Vi Institute of Technology, Sirunkundrum,  
Kanchepuram District, Tamilnadu, India

**ABSTRACT:** Now a days the life style of the people is different. People feel uncomfortable and time consuming for going crowded markets. So, E-Shopping is a boon as it saves lot of time. Online shopping is a process whereby consumers directly buy goods, services etc. To find out attractive products from online shopping marketplaces, the skyline query is a useful tool which offers more interesting and preferable choices for customers. The skyline query and its variants have been extensively investigated. However, to the best of our knowledge, they have not taken into account the requirements of customers in certain practical application scenarios. Recently, online shopping marketplaces usually hold some price promotion campaigns to attract customers and increase their purchase intention. Considering the requirements of customers in this practical application, we are concerned about product selection under price promotion.

**KEYWORDS:** *Skyline Query*

### I. INTRODUCTION

In our project, we are analyzing the data by using data mining. Data mining is the extraction of required data from large amount of raw data. The main concept in our project is separating the products; we will separate the desired products from huge number of products. Customer should get satisfied with the product quality and price.

We have proposed a skyline query which meets the satisfaction of the customer. The desired quality and price are taken into consideration and based on that the products are displayed. It returns the products that are not dominated by other products.

In our project, we formulated the COPC problem to retrieve optimal skyline product combinations that satisfy the customer's payment constraint and bring the maximum discount rate. To handle with the COPC problem, we have proposed an exact algorithm, and develop an incremental greedy algorithm to boost the performance. In addition to combinations of homogeneous products, we focused on the COPC problem over products of different categories. it is significant and interesting to compute optimal product combinations that meet different customer demands such as save or spend the most money under their budgets.

#### a.Data Mining:

Data Mining is defined as the procedure of extracting information from huge sets of data. In other words, we can say that data mining is mining knowledge from data.

There is a huge amount of data available in the Information Industry. This data is of no use until it is converted into useful information. It is necessary to analyze this huge amount of data and extract useful information from it.

Extraction of information is not the only process we need to perform; data mining also involves other processes such as Data Cleaning, Data Integration, Data Transformation, Data Mining, Pattern Evaluation and Data Presentation. Once all these processes are over, we would be able to use this information in many applications such as Fraud Detection, Market Analysis, Production Control, Science Exploration, etc.

The information or knowledge extracted so can be used for any of the following applications

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 3, March 2019

- Market Analysis
- Fraud Detection
- Customer Retention
- Production Control
- Science Exploration

Data Mining Applications:

Data mining is highly useful in the following domains

- Market Analysis and Management
- Corporate Analysis & Risk Management
- Fraud Detection

Apart from these, data mining can also be used in the areas of production control, customer retention, science exploration, sports, astrology, and Internet Web Surf-Aid.

Data mining involves six common classes of tasks:

Anomaly detection ( outlier / change / deviation detection ) – The identification of unusual data

- records, that might be interesting or data errors that require further investigation.
- Association rule learning (dependency modelling) – Searches for relationships between variables. For example, a supermarket might gather data on customer purchasing habits. Using association rule learning, the supermarket can determine which products are frequently bought together and use this information for marketing purposes. This is sometimes referred to as market basket analysis.
- Clustering – is the task of discovering groups and structures in the data that are in some way or another "similar", without using known structures in the data.
- Classification – is the task of generalizing known structure to apply to new data. For example, an e-mail program might attempt to classify an e-mail as "legitimate" or as "spam".
- Regression – attempts to find a function which models the data with the least error that is, for estimating the relationships among data or datasets.
- Summarization –providing a more compact representation of the data set, including visualization and report generation.

By using software to look for patterns in large batches of data, businesses can learn more about their customers to develop more effective marketing strategies, increase sales and decrease costs. Data mining depends on effective data collection, warehousing and computer processing.

## b. Projectuse in Domain:

In our project, we are analyzing the data by using data mining. Data mining is the extraction of required data from large amount of raw data. The main concept in our project is separating the products, we will separate the desired products from huge number of products. Customer should get satisfied with the product quality and price.

## c.Group Skyline Queries:

The skyline query aims to return the points that are not dominated by any other point [1]. However, most of the works about the skyline query just analyze individual points, and they are inappropriate to many applications that call for analysis of groups of different points. Motivated by this, group skyline queries are developed and paid growing attention. In most of the group skyline queries, optimal groups are computed by the dominance relationship between corresponding aggregate-based points of different groups.

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 3, March 2019

## II. RELATED WORK

Su et al. [3] formulated top k combinatorial skyline query (kCSQ). It returns those combinatorial skyline tuples whose aggregate values for a certain attribute are maximum. Since only the first k combinations are required, the k-CSQ query process can be simplified [4]. Chung et al. [4] extended the traditional skyline queries and formulated a combinatorial skyline query, namely CSQ, which is to find the outstanding skyline combinations. Im et al. [5] studied the group skyline query which is based on the dominance relationship between the groups of the same size. The dominance relationship is checked according to the aggregate values of attributes. Magnani et al. [6] introduced aggregate skylines, where the skyline works as a filtering predicate on sets of records.

The aggregate skyline queries merge the functionalities of two basic database operators, skyline and group by. Zhang et al. [7] focused on a novel problem of groups of k tuples, which are not dominated by any other group of equal size, based on aggregate-based group dominance relationship. They also identified two anti-monotonic properties to filter out candidate groups. Wu et al. [2] researched the work which is similar to the group skyline computation. They focused on a problem of creating competitive products which are not dominated by the products in the existing market. Here each new product is generated by combining products from different source tables. To reduce the search space, it only combines the skyline products from source tables to generate new products.

In addition, they also presented an approach which divides similar products into groups and processes them as a whole. In the group skyline queries aforementioned, the aggregate values of corresponding attributes are taken into account when checking the dominance relationship between different groups. However, it is difficult for customers to specify an appropriate aggregate function.

Moreover, it overlooks many significant groups that may contain nonskyline points [8]. Motivated by the problems which the group skyline queries face, Liu et al. [8] presented a Pareto group-based skyline (G-skyline) query which retrieves Gskyline groups.

These G-skyline groups are not g-dominated by any other group of the same size. They presented a directed skyline graph which captures the dominant relationship of tuples within the first k skyline layers. Furthermore, two heuristic algorithms are designed to solve the G-skyline query effectively.

Recently, Yu et al. [9] defined the Multiple Skyline Layers, and presented two fast algorithms to compute the G-skylines due to the observation that skyline points contribute more to skyline groups compared to non-skyline points. The above approaches for group skyline queries are inappropriate for our problem. This is since they filter out the optimal groups by the dominance relationship, but the skyline product combinations in our COPC problem are measured by the maximum discount rate.

This query does not necessarily contain skyline points but has the advantages of both ranking queries and skyline queries, which are with the control on the size of the answer set and without users' efforts to specify ranking functions. Chan et al. [15] formulated a top k ranking problem which retrieves points appearing frequently in the subspace skylines. Wan et al. [14] identified the problem of finding top-k profitable products.

Given a set of packages in the existing market and a set of potential new packages, they wanted to select k new packages such that the sum of the profits of the selected packages is maximized and each selected package is not dominated by any package in the existing market and any selected new package. Lin et al. [13] proposed a k-most demanding products (k-MDP) discovering problem that helps the company to select k products from the candidate products with the maximum expected number of the total customers. In [20], we developed a top k favorite probabilistic product query. It is utilized to select k products which can meet the needs of a customer set at the maximum level. The representative skyline queries aforementioned only highlight some features which are stability, scale invariance, diversification of the results, and partial knowledge of the record scoring function [16]. Magnani et al. [16] focused on the representative skyline queries in terms of both the significance and diversity of results. This representative 1041-4347 (c) 2018 IEEE. In [22], Chen et al. studied a constrained skyline query. It first filters out the points which do not satisfy range constraints of each attributes, and then processes the skyline query over the remaining points. To our best knowledge, this is the first attempt to research product selection under price promotion.

The COPC problem aims to find optimal skyline product combinations whose actual payments are not beyond the customer's willingness to pay. Moreover, these optimal skyline product combinations could bring the maximum

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 3, March 2019

discount rate. Noticeably, the methods for the skyline queries under constraints cannot be applied to process our problem straightly.

**b.Skyline Queries under Constraints---** Skyline query is a useful tool to find out attractive products which offer more interesting and preferable choices for customers. However, the size of the skyline query results cannot be controlled flexibly. Accordingly, many research efforts have been devoted to contend with this problem. The existing approaches to address this problem are developed to identify  $k$  representative skylines which have the maximum dominant capacity or the maximum diversification. Lu et al. [10] were concerned about the case when the actual cardinality of skyline results is less than the desired result cardinality  $k$ . They proposed a new approach, namely skyline ordering, which forms a skyline-based partitioning of a given dataset. Then they applied a set-wide maximization technique, which is used to find an object set dominating the largest number of points, to process each partition. Bai et al. [19] focused on how to choose  $k$  representative skylines over data streams. Gao et al. [18] formulated the most desirable skyline object query which reports the most preferable  $k$  skylines based on a new ranking criterion. Lin et al. [12] studied the problem of selecting  $k$  skyline points such that the number of points, which are dominated by at least one of these  $k$  skyline points, is maximized.

Papadias et al. [11] proposed the  $K$ -dominating query which retrieves  $K$  points dominating the largest number of other points. This query does not necessarily contain skyline points but has the advantages of both ranking queries and skyline queries, which are with the control on the size of the answer set and without users' efforts to specify ranking functions. Chan et al. [15] formulated a top  $k$  ranking problem which retrieves points appearing frequently in the subspace skylines. Wan et al. [14] identified the problem of finding top- $k$  profitable products. Given a set of packages in the existing market and a set of potential new packages, they wanted to select  $k$  new packages such that the sum of the profits of the selected packages is maximized and each selected package is not dominated by any package in the existing market and any selected new package. Lin et al. [13] proposed a  $k$ -most demanding products ( $k$ -MDP) discovering problem that helps the company to select  $k$  products from the candidate products with the maximum expected number of the total customers. In [20], we developed a top  $k$  favorite probabilistic product query. It is utilized to select  $k$  products which can meet the needs of a customer set at the maximum level. The representative skyline queries aforementioned only highlight some features which are stability, scale invariance, diversification of the results, and partial knowledge of the record scoring function [16]. Magnani et al. [16] focused on the representative skyline queries in terms of both the significance and diversity of results. This representative 1041-4347 (c) 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. skyline query can satisfy all the features above. Moreover, there are also some other approaches to avoid returning many skyline query results. In [22], Chen et al. studied a constrained skyline query. It first filters out the points which do not satisfy range constraints of each attributes, and then processes the skyline query over the remaining points. To our best knowledge, this is the first attempt to research product selection under price promotion. The COPC problem aims to find optimal skyline product combinations whose actual payments are not beyond the customer's willingness to pay. Moreover, these optimal skyline product combinations could bring the maximum discount rate. Noticeably, the methods for the skyline queries under constraints cannot be applied to process our problem straightly.

## III. PROPOSED WORK

We have proposed to find skyline product combinations which meet a customer's payment willingness and bring the maximum discount rate. A skyline query returns the objects that cannot be dominated by any other objects, which leads to customer satisfaction. Besides, we design an algorithm, which has guarantee about the accuracy of the results.

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 3, March 2019



Fig 1.Overall Process Design

The above figure represents the overall design of our proposed framework.

## IV.IMPLEMENTATION

### 1. AUTHENTICATION:

All the user and the seller account was authenticated by the Admin. After the Approval of the Seller product by the Admin, the products are displayed to the users with the valid price. The Admin has to approve the payment of the product. The Admin can view all the Seller and the User details. The Admin can compose mail and see seller mails from Inbox.

### 2. PRODUCT MANAGEMENT:

The product management is performed by the Seller. The seller can add products and view all the products. The Seller can view mails from the Admin and the User. The products which are added by the Seller will be displayed to the user/customer only after the approval of those products by the Admin. The seller can also view the Purchased products and bargained products. The payment for the product is received by the Seller.

### 3. PURCHASE MODULE:

Online market where various goods are bought by the registered user. The user can choose all the products approved by the Admin and select the product to purchase. The Quality, Comment and the Ratings for the product are displayed to the User. The User place their Bargain request for the selected product. The User also add comment and the Ratings for the product. The User can check mails from the Seller and also compose mail to the Seller.

### 4. IMPLEMENTATION OF SKYLINE QUERY:

The request from user are processed by the skyline query. Apply skyline query to the user bargain request. It gets the output from the admin and Send the output to the seller and the Customer receives output from seller.

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 3, March 2019

## V. CONCLUSION

Online Shopping is a web-based project which is made for remote-shopping through Internet. As the technology is being advanced the way of life is changing accordance. Now a day's we can place the order for anything from our home. There is no need to go shop of the things we want. The order can be placed online through internet.

The payment and the confirmation of purchasing can be done .Now we can see that how the days have been changed with time. People had to stand in rows to wait there terms to buy a particular thing from a popular shop. But what is happening now a day's, we can extremely surprise that those things can be available on the door-step in few hours as per our choice. In future we will try to make this website which work so flexible and beneficial for the customer and also try to make smooth service.

## REFERENCES

- [1] S. Borzsonyi, D. Kossmann, and K. Stocker, "The skyline operator," in Proc. Int'l Conf. Data Eng. (ICDE), pp. 421–430, 2001. [20] X. Zhou, K. Li, G. Xiao, Y. Zhou, and K. Li, "Top k favorite probabilistic products queries," IEEE Trans. on Knowl. Data Eng., pp. 2808–2821, 2016.
- [2] Y.-C. Chung, I.-F. Su, and C. Lee, "Efficient computation of combinatorial skyline queries," Information Systems, vol. 38, no. 3, pp. 369–387, 2013.
- [3] H. Im and S. Park, "Group skyline computation," Information Sciences, vol. 188, pp. 151–169, 2012.
- [4] J. Liu, L. Xiong, J. Pei, J. Luo, and H. Zhang, "Finding pareto optimal groups: Group-based skyline," Proc. of the VLDB Endowment, vol. 8, no. 13, 2015.
- [5] W. Yu, Z. Qin, J. Liu, L. Xiong, X. Chen, and H. Zhang, "Fast algorithms for pareto optimal group-based skyline," in Proc. Int. Conf. on Information and Knowledge Management, pp. 417–426, 2017.
- [6] D. Papadias, Y. Tao, G. Fu, and B. Seeger, "Progressive skyline computation in database systems," ACM Transactions on Database Systems (TODS), vol. 30, no. 1, pp. 41–82, 2005.
- [7] X. Lin, Y. Yuan, Q. Zhang, and Y. Zhang, "Selecting stars: The k most representative skyline operator," in Proc. 23th Int'l Conf. Data Eng. (ICDE), pp. 86–95, IEEE, 2007. [13] C.-Y. Lin, J.-L. Koh, and A. L. Chen, "Determining k-most demanding products with maximum expected number of total customers," IEEE Trans. on Knowl. Data Eng., vol. 25, no. 8, pp. 1732–1747, 2013.
- [8] Q. Wan, R.-W. Wong, and Y. Peng, "Finding top-k profitable products," in Proc. 27th Int'l Conf. Data Eng. (ICDE), pp. 1055–1066, IEEE, 2011.
- [9] G. Xiao, K. Li, X. Zhou, and K. Li, "Efficient monochromatic and bichromatic probabilistic reverse top-k query processing for uncertain big data," Journal of Computer & System Sciences, vol. 89, 2016.
- [10] Y. Gao, Q. Liu, L. Chen, G. Chen, and Q. Li, "Efficient algorithms for finding the most desirable skyline objects," Knowledge-Based Systems, vol. 89, no. C, pp. 250–264, 2015. [11] M. Bai, J. Xin, G. Wang, and L. Zhang, "Discovering the k representative skyline over a sliding window," IEEE Trans. on Knowl. Data Eng., vol. 28, no. 8, pp. 2041–2056, 2016