



# **iRecommender**

For a Better E-Commerce Experience

CDAP Proposal Presentation  
Group -17-035



# What is E-Commerce

- ✓ **E-commerce** is a transaction of buying or selling online

## Why E-Commerce

- ✓ **E-commerce** easy to use than visiting a physical store.
- ✓ E-commerce is easy to manage than managing a physical store



# Research Problems

1. Product Suggestions based on search Results are inconsistent.
2. Not updating Suggestions with information outside the website.
3. Time waste due to inefficient product suggestions
4. Reduction of users impressions on specific E-Commerce websites.
5. Struggle on suggesting products to increase sales by the e-commerce owners.



# E-Commerce Recommendation Systems

## **1. Content based recommendation Engines.**

Based on the description of the item and the profile of the user's preference.

## **2. Collaborative Recommendation Engines**

Collaborative filtering is a method of making automatic predictions (filtering) about the interests of a user by collecting preferences or taste information from many users (collaborating)



# Drawbacks of existing Systems

- 1. Cold Start Problem**
- 2. Database Delay**
- 3. No scalability of the system**
- 4. Inefficient data cluster management**
- 5. Usage of only one type of recommendation technology**

# Evaluation of existing Recommendation Engines

	Netflix	eBay	Amazon	Ali Express	Walmart
Cold Start Problem	✓	✓	✓	✓	✓
Database delay	✓	X	X	X	X
No scalability	X	X	X	X	X
Inefficient cluster management	✓	X	X	X	X
Usage of only one type of recommendation technology	✓	✓	✓	✓	✓



# What Researchers have done so far

- Two Recommendation methodology approaches
  - ✓ Collaborative based recommendation
  - ✓ content based recommendation
- Most of the Researchers have used one recommendation approach.
- Most of the time that recommendations are not feasible and accurate.
- Tried to make a solution for overcome the “Cold Start” problem.

# Cont.

- Researchers suggest deep learning, content based recommendation, and Hybrid Recommendation (Collaborative based+ Content based) for overcome the Cold start problem
- Response time is very high in existing systems because most of the recommendation engines use relational databases.
- So our main targets are to
  - Overcome the cold start problem
  - Make the recommendations in high accuracy
  - Overcome the database delay( quick response time )
  - Make the System Scalable



# Our Solution

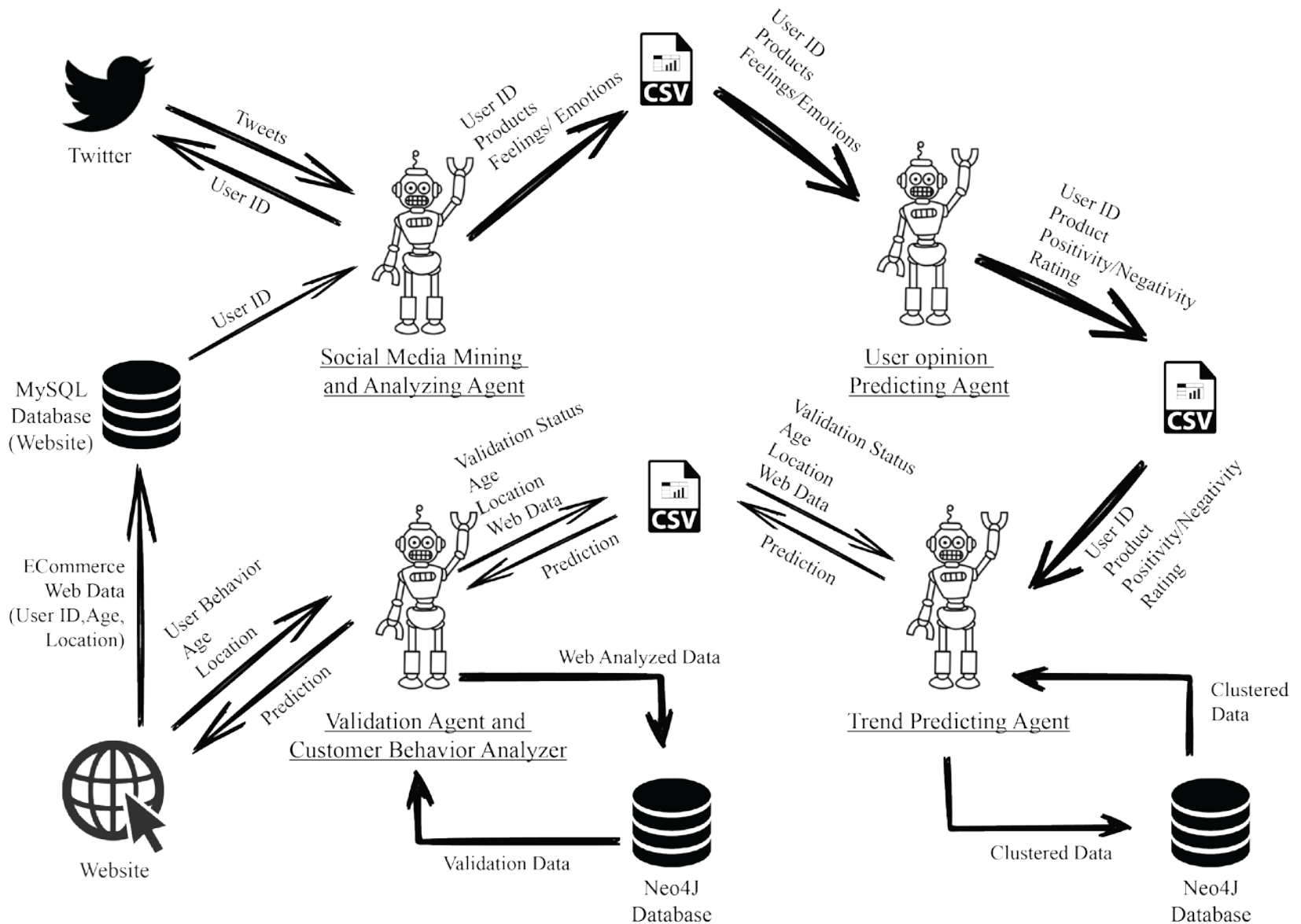


**iRecommender**  
For a Better E-Commerce Experience

**Social media data involved hybrid  
and graph based database solution**



# iRecommender





# Benefits for the Users Of the System

## **Business Perspective:-**

1. Increase sales of the e-commerce web site
2. Increase loyalty of the Customers
3. Overcome the struggle of product suggestion

## **Customer Perspective**

1. Customer no need to have any special technical knowledge to search
2. Save the Time

# **Sub Module 1:- Social media mining and Analyzing Agent**

main role is collect the data as a inputs to predict customers behavior. To this have to analyze tweets in each customer.

## **main Responsibilities**

- i. Analysis each customer:s social media**
- ii. Remove noise words of tweets**
- iii. Identify the products name if customer have mentioned web site:s selling products.**
- iv. Identify the emotional words which can get customer:s opinion about that products.**
- v. Finally Pass the Data to user opinion predictor agent**

- ***Social Media Analyzer***

Twitter has attracted much interest the last few years from researchers, because Twitter has provided a public API with many resources to developer perspectives to do some research.

- ***Noise word removal***

Sentiment classification over Twitter is usually affected by the noisy nature of tweets data.

A popular procedure to reduce the noise of textual data is to remove noise words by using pre-compiled noise words lists or more sophisticated methods for dynamic noise words identification.

- By using pre-define noise words list.
- Removing single letters / punctuation marks and other symbols.
- Removing words with low inverse document frequency

- ***Identifying the products name and emotional words***

to identify the products names by looking tweets words which are selling e-commerce site.

develop products dictionary by using e-commerce web site's database data.

which are the products customer have mentioned by using lexical analyze.

- ***Identify the emotional words***

- use algorithm is the emotional dictionary of the 'Linguistic Inquiry and Word Count'
- contains a broad dictionary list combined with emotional categories

➤ Final outcome

Tweet ' - I like lather wristwatch

Outcome' - Like , Wristwatch



# **Sub Module 2 :- User Opinion Predicting Agent**

## **What is an Opinion:**

- The simple meaning of the Opinion is a personal belief or judgment that is not founded on proof or certainty.
- It is a personal view, attitude, or appraisal.
- It is a formal expression of a professional judgment.
- Generally, an Opinion can be a Negative like Opinion, Positive like Opinion or Neutral like Opinion.

# What is an Opinion to a Machine?

- It is a "quintuple", an object made up of 5 different things

$$(o_j, f_{jk}, so_{ijkl}, h_i, t_l)$$

- **$O_j$**  - The thing in question (i.e. product) / Target entity.
- **$F_{jk}$**  - A feature/aspect of  $o_j$ .
- **$SO_{ijkl}$**  - The sentiment value of the opinion of the opinion holder  $h_i$  on feature  $f_{jk}$  of object  $o_j$  at time  $t_l$ .  $so_{ijkl}$  is Positive, Negative or Neutral.
- **$H_i$**  - Opinion holder.
- **$T_i$**  - The time when Opinion is expressed

✓ Machine has to be identified these 5 elements.



## The Main Process

- The main Objective of this section goes to extract the Negativity or Positivity of the feelings.
- There are 3 main sub parts.
  - Get set of data.
  - Execute User Opinion Predict Algorithm for Opinion Mining.
  - Store the output result.

## **What is the User Opinion Predictor Algorithm:**

- This can be done by using Sentiment Analysis.
- Basic Sentiment Analysis algorithms use Natural Language Processing (NLP) to classify words as positive, neutral, or negative.
- Keyword spotting is the simplest technique leveraged by sentiment analysis algorithms.
- Algorithms vary in the way they score the words to decide whether they indicate overall positive or negative sentiment.
- Different algorithms have different libraries of words and phrases which they score as positive, negative, and neutral.

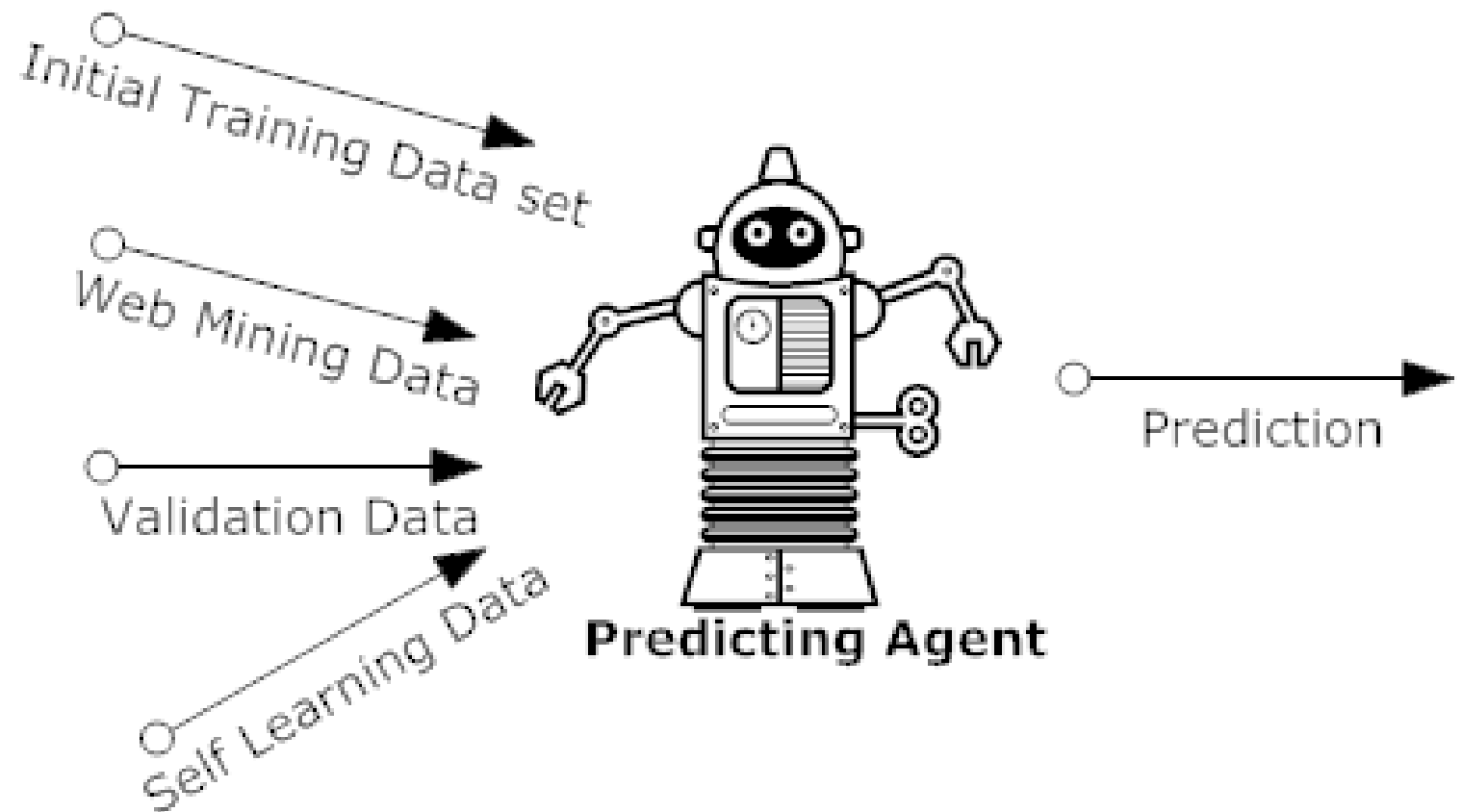
## Cont..

- Output of this User Opinion Predictor Algorithm is Positivity Negativity or Neutrality according to the opinions.
- This will order as below
  - 0 - Very Negative
  - 1 - Negative
  - 2 - Neutral
  - 3 - Positive
  - 4 - Very Positive
- Under the User ID, these negativity, positivity or neutrality is stored in .csv file.
- For each user, there is a unique .csv file which under naming the User ID.

# Sub Module 3- Trend Predicting Agent

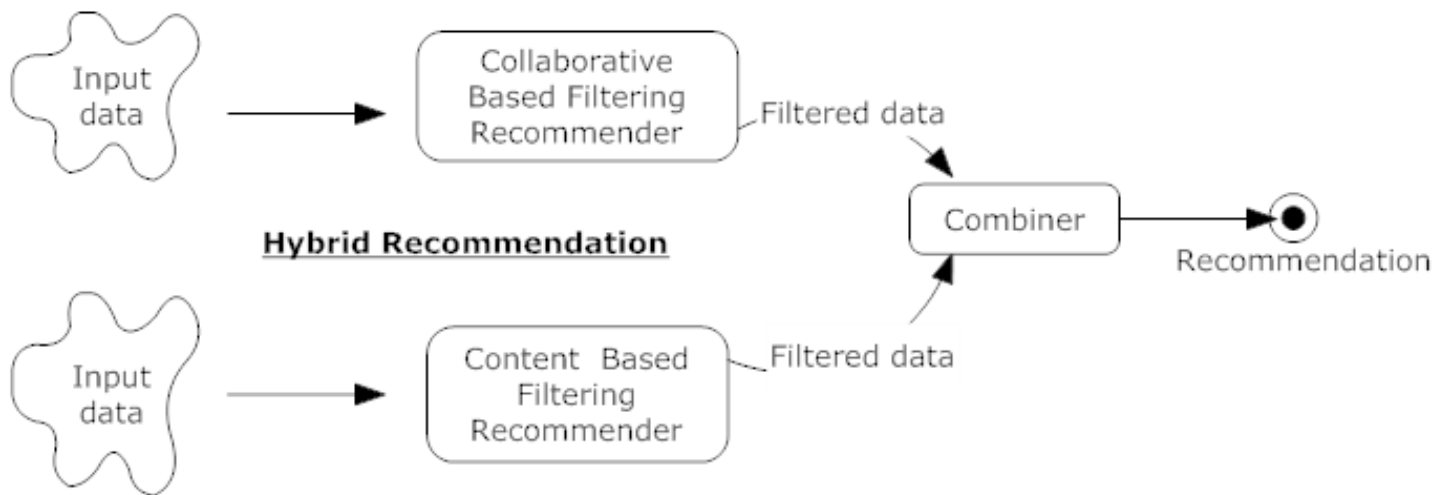
**Who is the Trend predicting agent:**

This component is for to predict the taste of the customer



# How to Fill the Existing Research Gap

Overcome the "Cold Start" Problem: -A Hybrid filtering system



## Machine Learning Algorithms Suppose to Use

- Cosine Similarity Algorithm
- K Nearest Neighbors Algorithm

**Overcome the database delay ,System Scalability and the response time Problem :**

**Use a graph database with Spark**



### **Why And What Is the Spark**

Apache Spark' is a fast and general engine for large-scale data processing

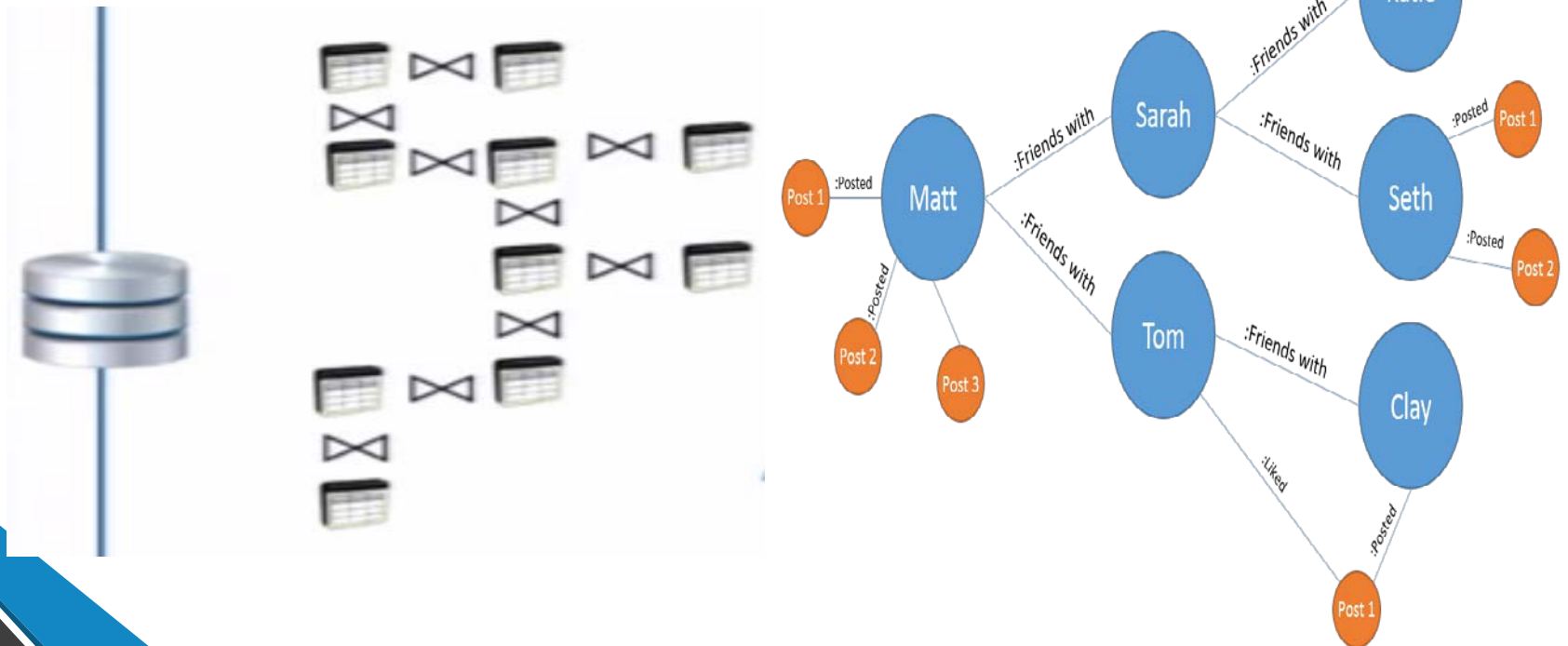
### **Speed**

Run programs up to 100x faster than Hadoop in memory, or 10x faster on disk

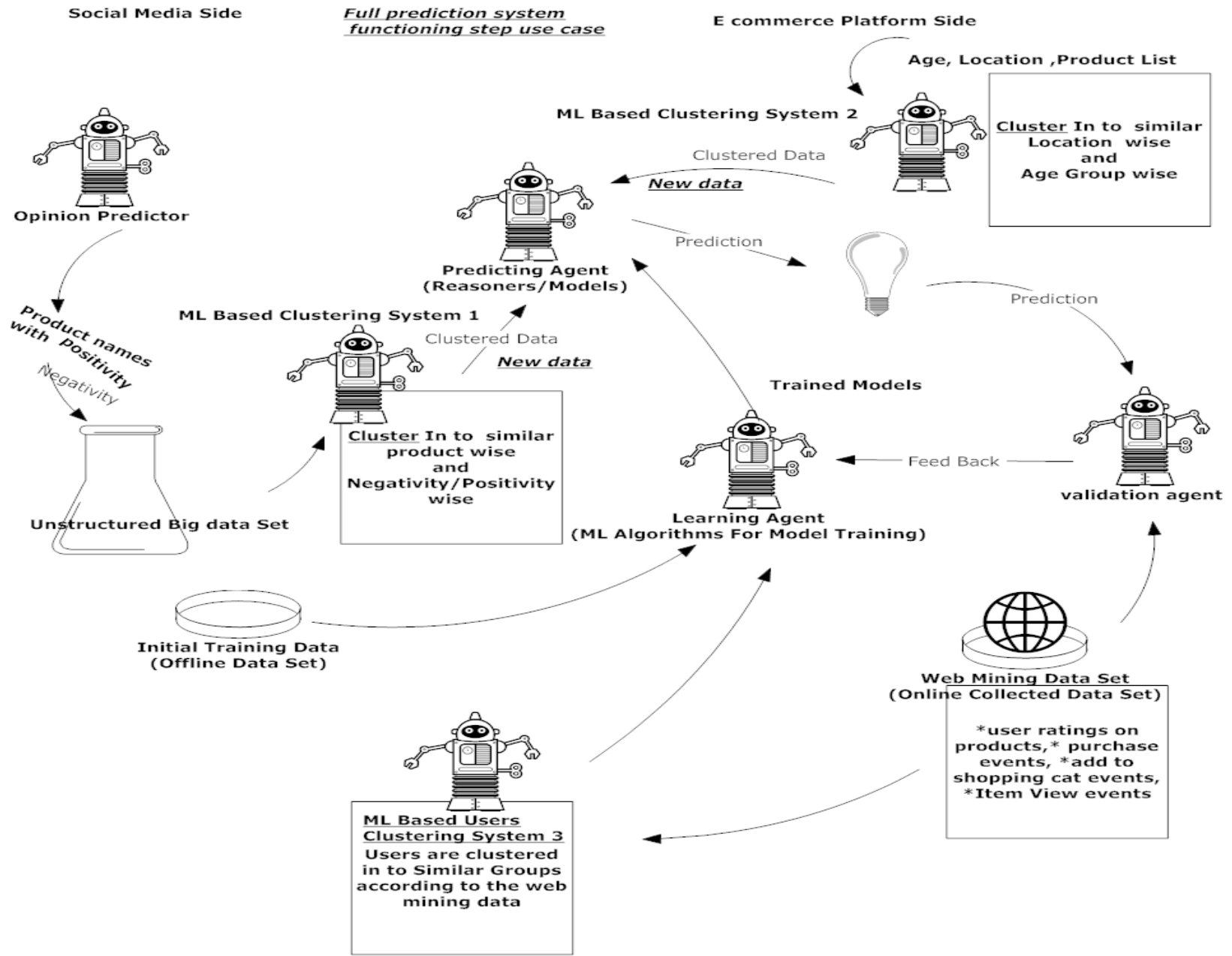
# Why And What Is the Neo4j

A Hi Speed graph NoSQL database management system with Create, Read, Update and Delete 'CRUD' operations working on a graph data model

## RDBMS Vs Graph Databases

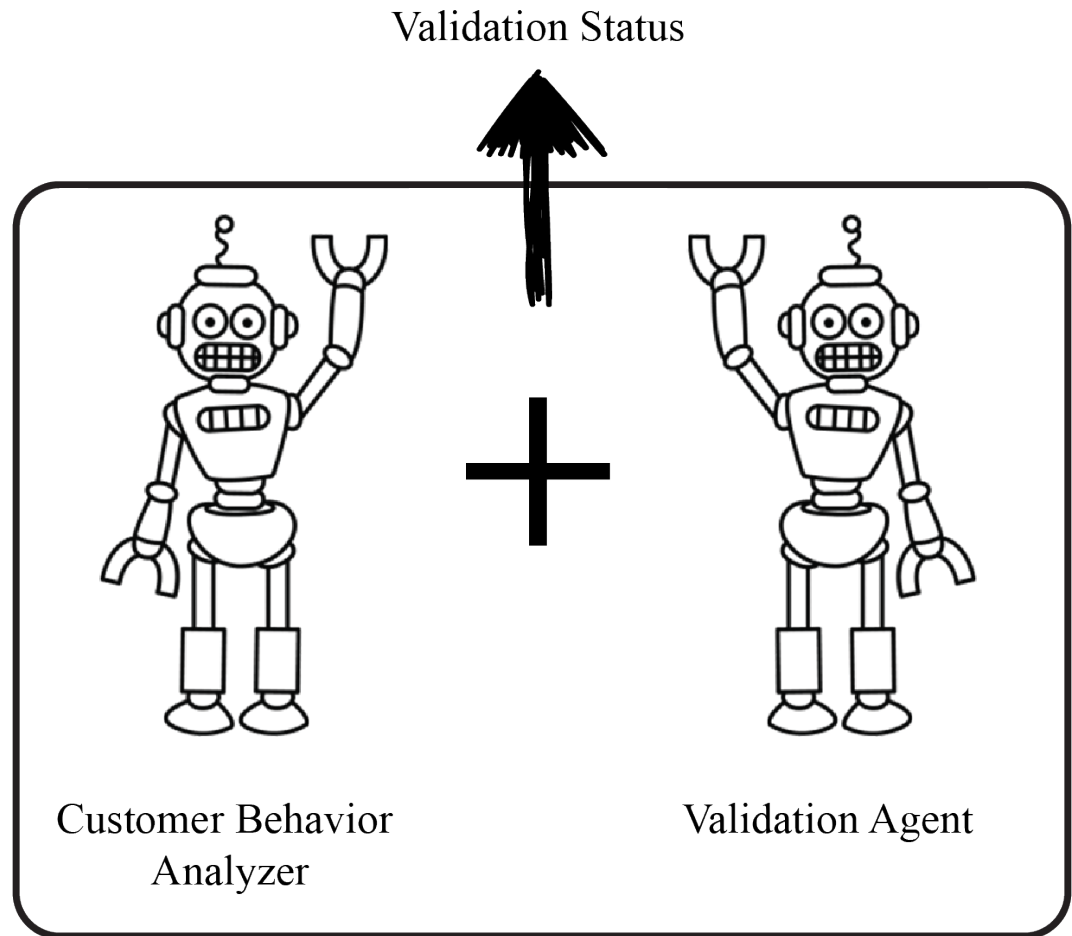


# How does the Prediction System works

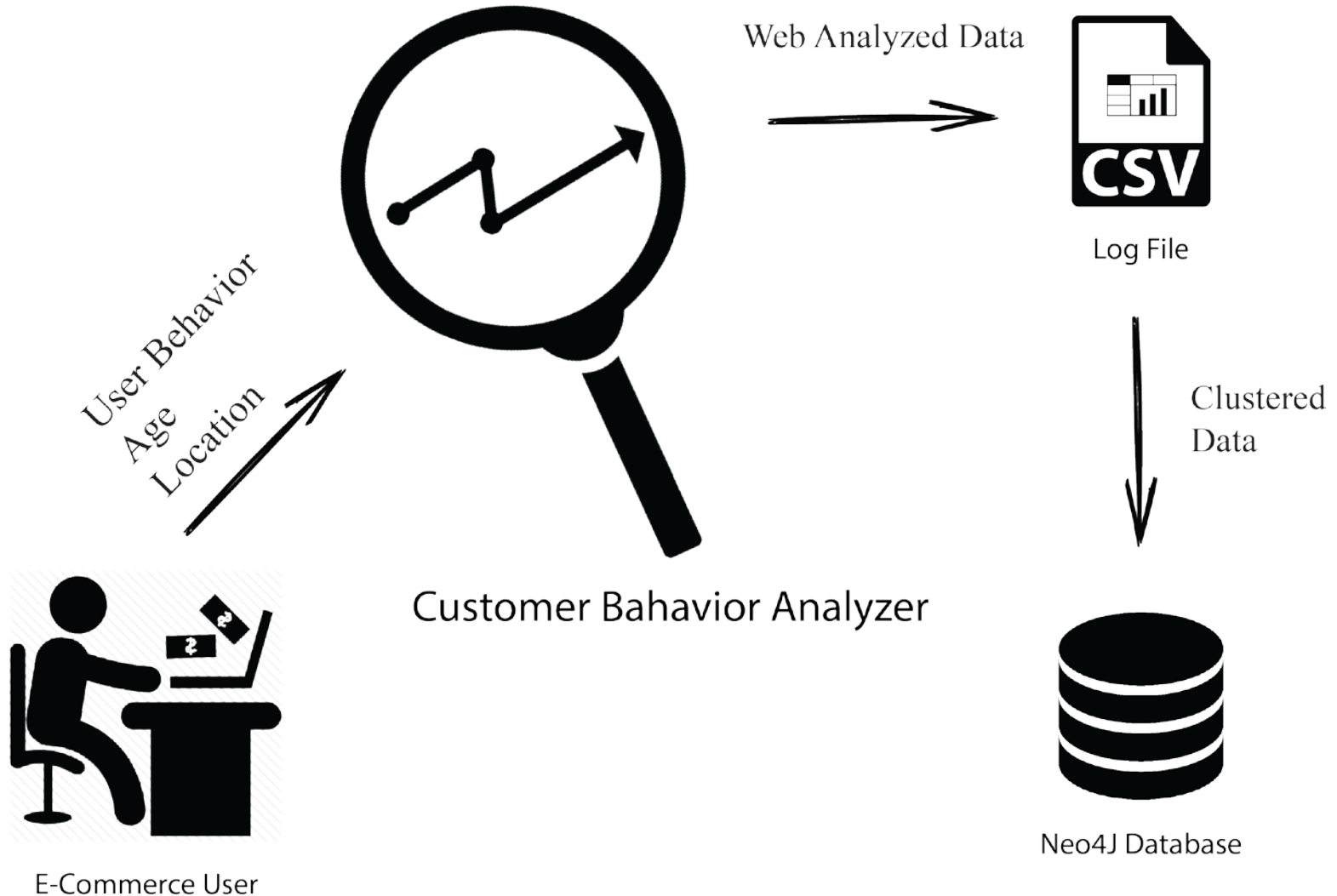




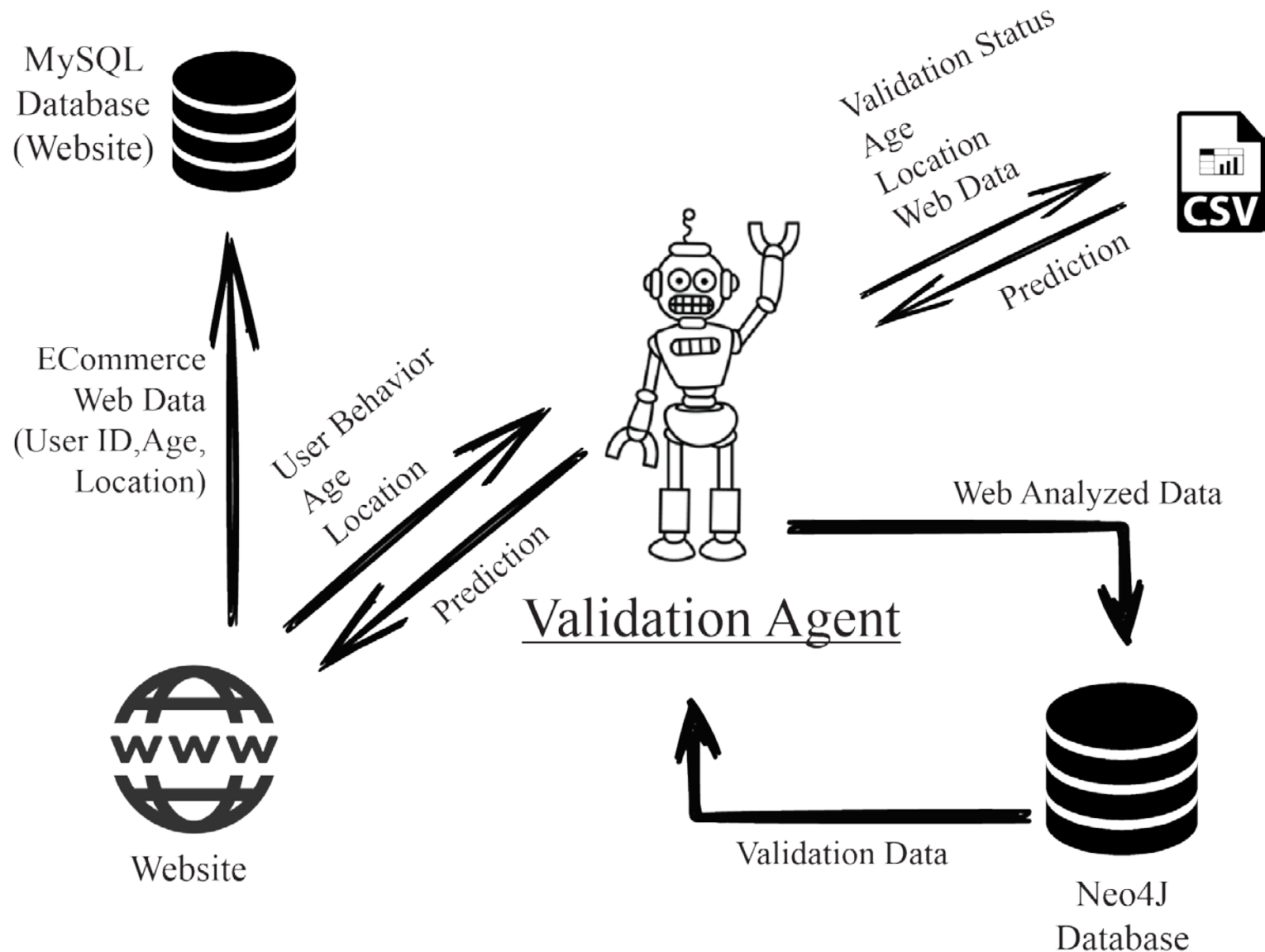
# Sub Module 4:- Validation Agent and Customer behavior Analyzer



# Customer behavior Analyzer



# Validation Agent



# Commercial Value Of the product

- By filling the existing GAP in the Recommendation engines Category We are trying to Make Best and High accurate recommendation engine for the e commerce



Thank you...!