# WAREHOUSE PROJECT

By : Omar Alomari and Osamah Kfaween and Yousef Shaban

# SOME OF DATA MINING ISSUES

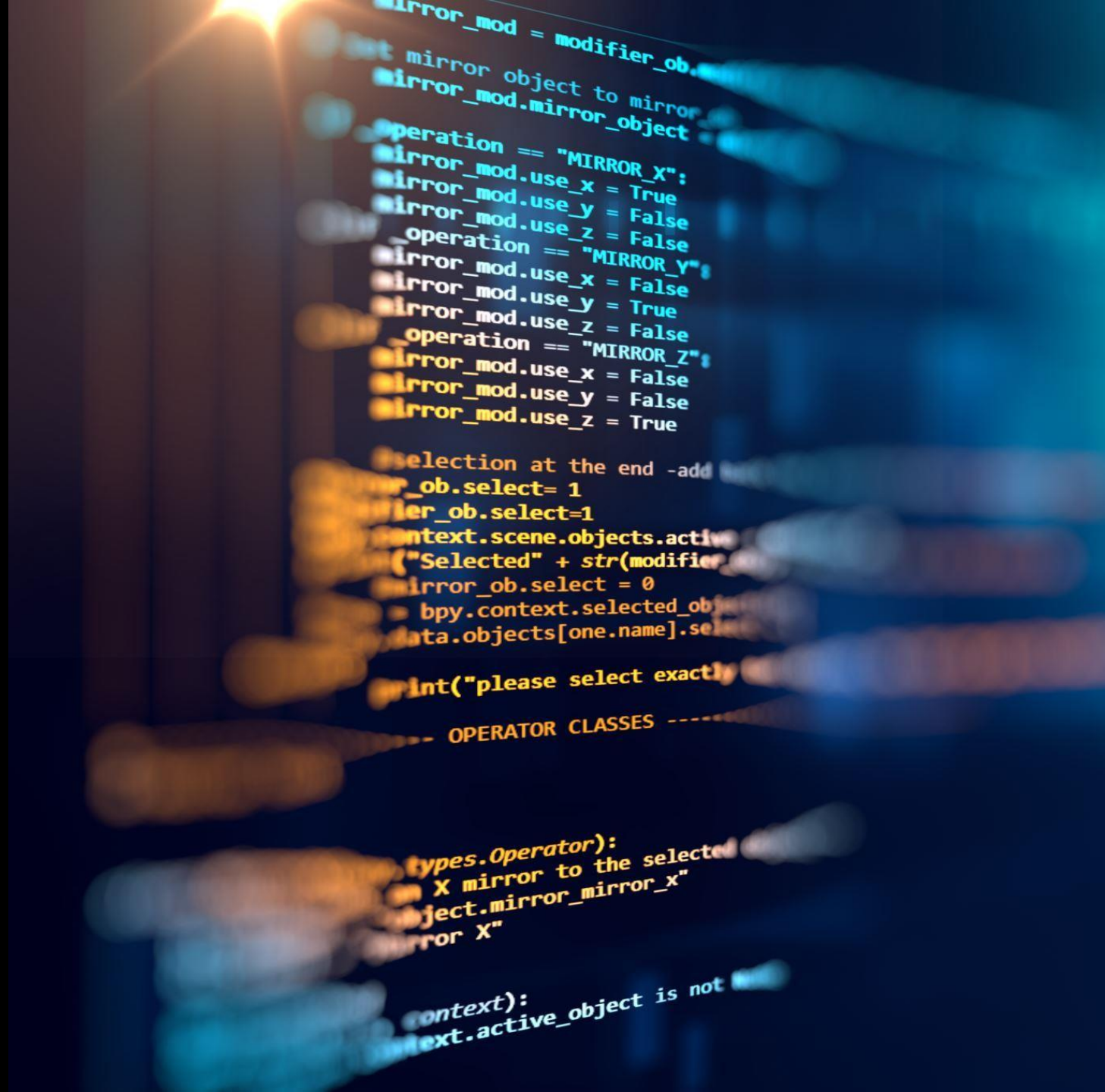
- Security and Social Challenges
- Noisy and Incomplete Data
- Distributed Data
- Complex Data
- Performance
- Scalability and Efficiency of the Algorithms
- Improvement of Mining Algorithms
- Incorporation of Background Knowledge
- Data Visualization
- Data Privacy and Security

# NOISY AND INCOMPLETE DATA

Data Mining is the way toward obtaining information from huge volumes of data. This present reality information is noisy, incomplete, and heterogeneous. Data in huge amounts regularly will be unreliable or inaccurate. These issues could be because of human mistakes blunders or errors in the instruments that measure the data.

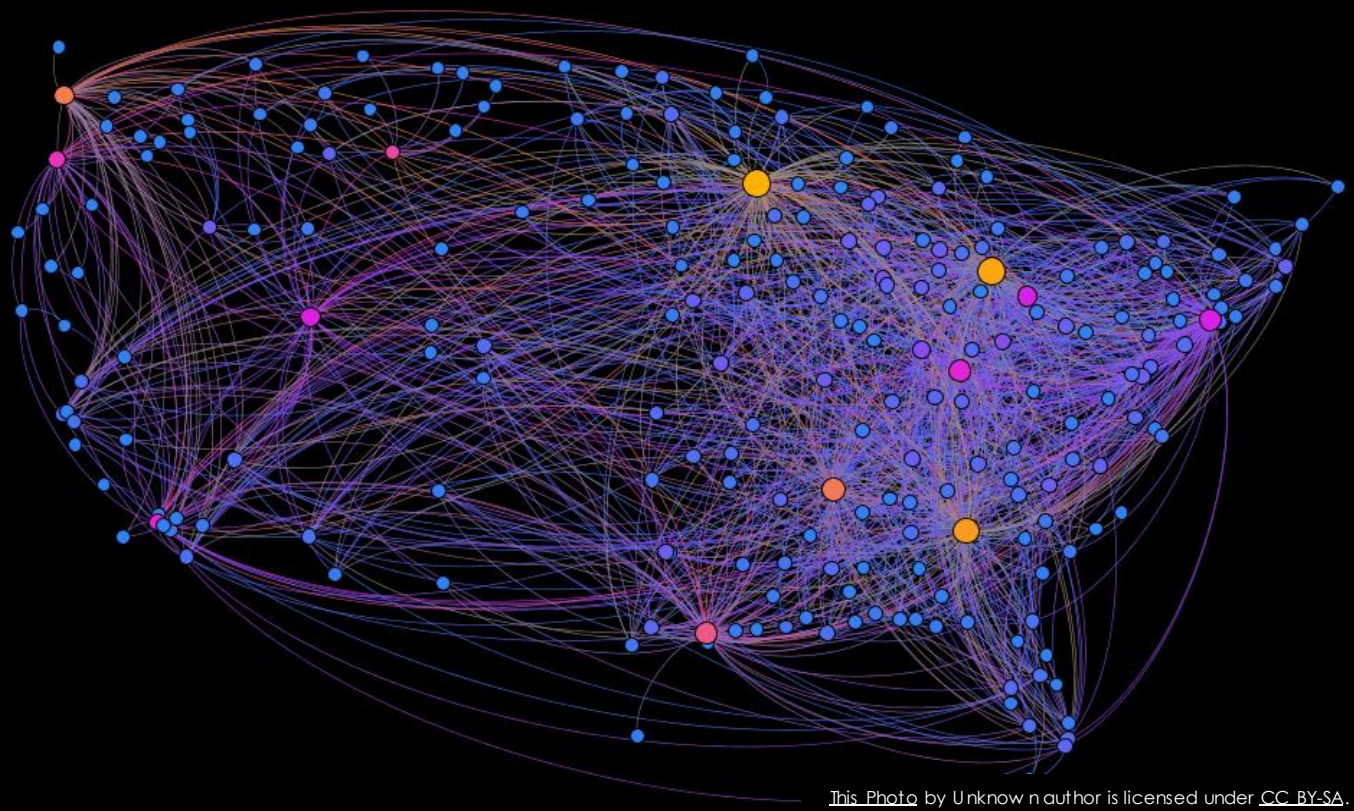
# COMPLEX DATA

True data is truly heterogeneous, and it very well may be media data, including natural language text, time series, spatial data, temporal data, complex data, audio or video, images, etc. It is truly hard to deal with these various types of data and concentrate on the necessary information. More often than not, new apparatuses and systems would need to be created to separate important information.



# DATA VISUALIZATION

Data visualization is a vital cycle in data mining since it is the foremost interaction that shows the output in a respectable way to the client. The information extricated ought to pass on the specific significance of what it really plans to pass on. However, ordinarily, it is truly hard to address the information in a precise and straightforward manner to the end-user. The output information and input data being very effective, successful, and complex data perception methods should be applied to make it fruitful.

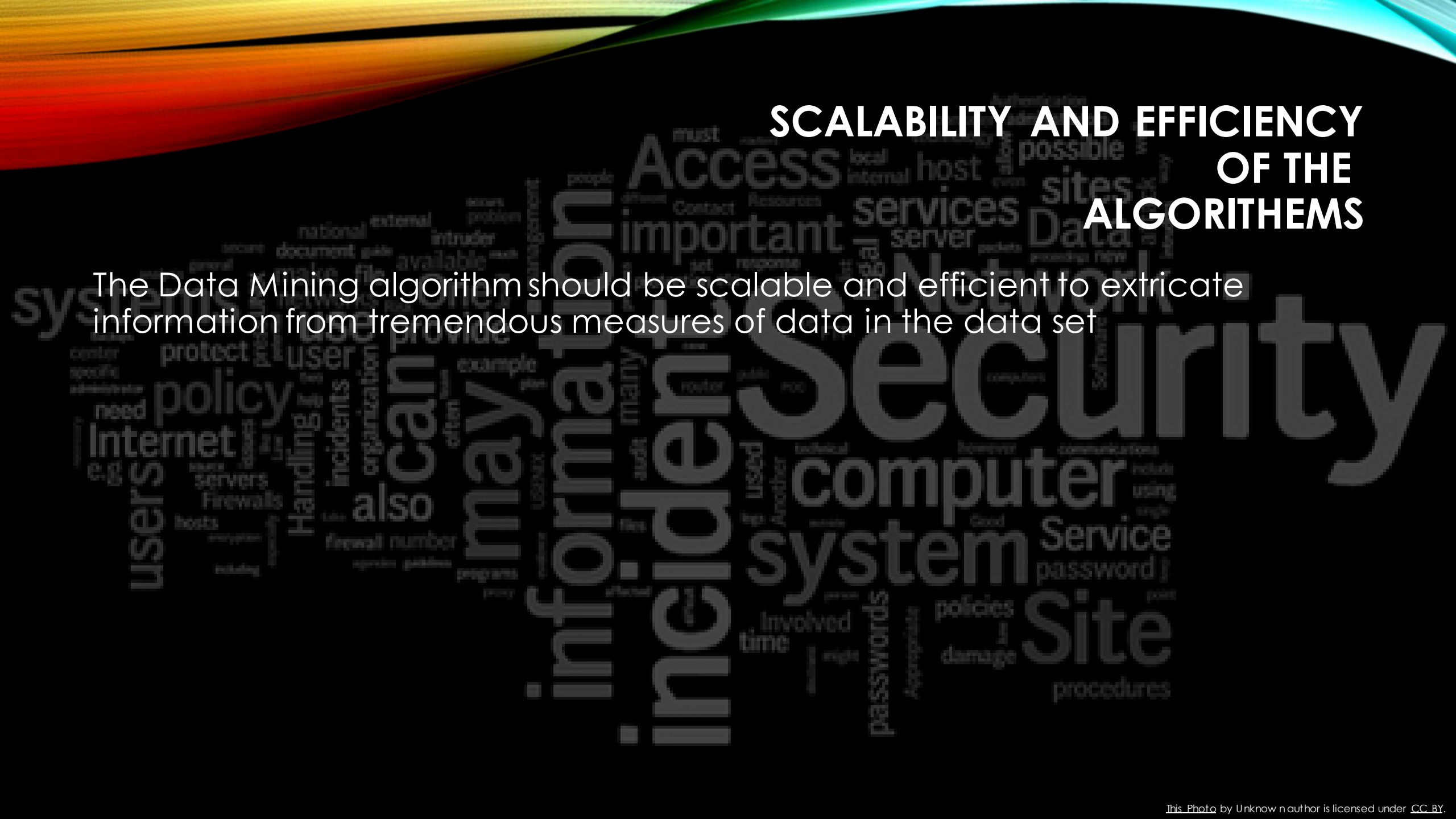


# DATA PRIVACY AND SECURITY

Data mining typically prompts significant issues regarding governance, privacy, and data security. For instance, when a retailer investigates the purchase details, it uncovers information about purchasing propensities and choices of customers without their authorization.

# SCALABILITY AND EFFICIENCY OF THE ALGORITHEMS

The Data Mining algorithm should be scalable and efficient to extricate information from tremendous measures of data in the data set



# PERFORMANCE

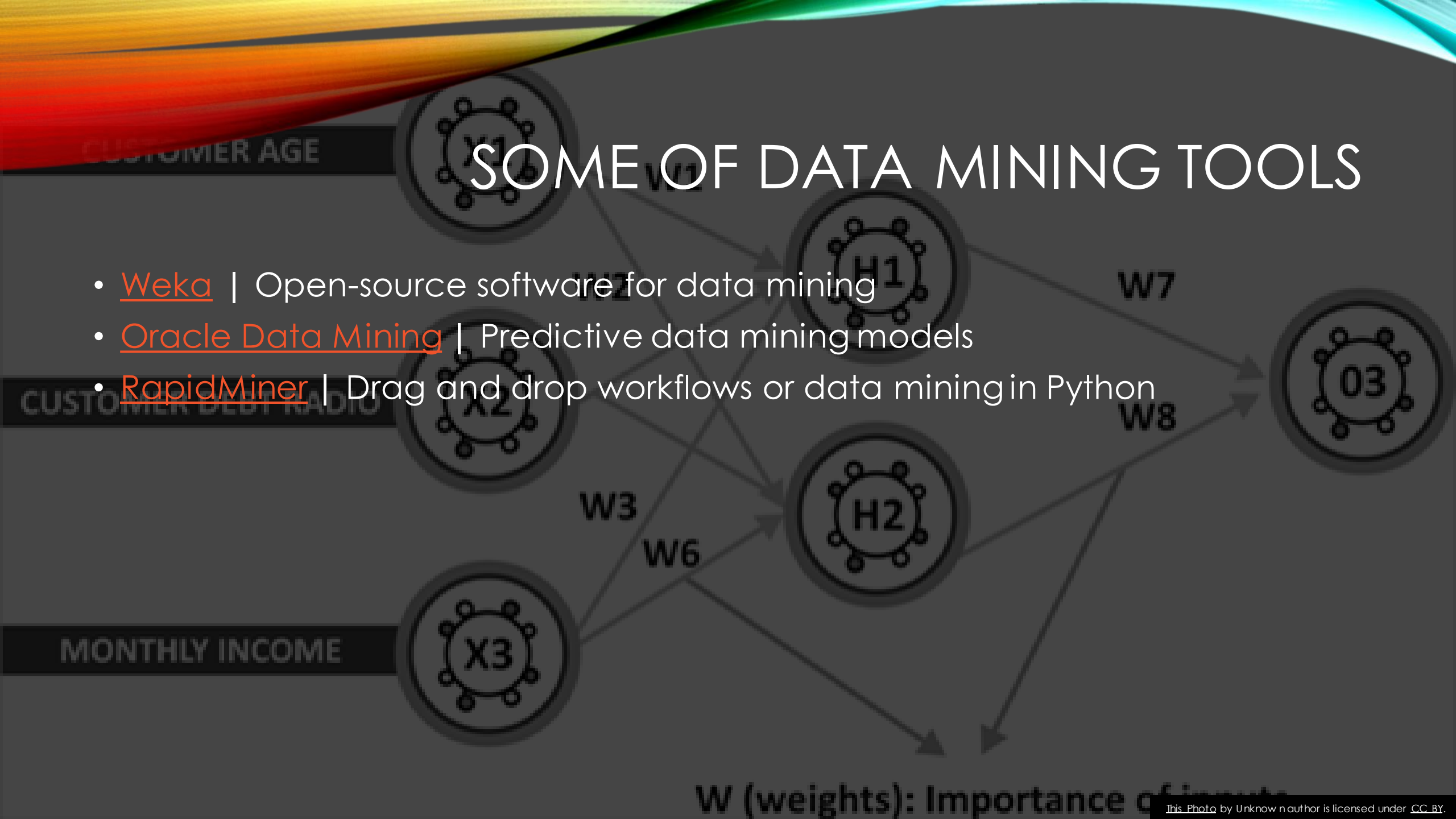
The presentation of the data mining framework basically relies upon the productivity of techniques and algorithms utilized. On the off chance that the techniques and algorithms planned are not sufficient; at that point, it will influence the presentation of the data mining measure unfavourably

# DISTRIBUTED DATA

True data is normally put away on various stages in distributed processing conditions. It very well may be on the internet, individual systems, or even on databases. It is essentially hard to carry all the data to a unified data archive principally because of technical and organizational reasons

# SOME OF DATA MINING TOOLS

- Weka | Open-source software for data mining
- Oracle Data Mining | Predictive data mining models
- RapidMiner | Drag and drop workflows or data mining in Python




# Weka

WEKA IS AN OPEN-SOURCE MACHINE LEARNING SOFTWARE WITH A VAST COLLECTION OF ALGORITHMS FOR DATA MINING. IT WAS DEVELOPED BY THE UNIVERSITY OF WAIKATO, IN NEW ZEALAND, AND IT'S WRITTEN IN JAVASCRIPT.


# Oracle data mining

ORACLE DATA MINING IS A COMPONENT OF ORACLE ADVANCED ANALYTICS THAT ENABLES DATA ANALYSTS TO BUILD AND IMPLEMENT PREDICTIVE MODELS. IT CONTAINS SEVERAL DATA MINING ALGORITHMS FOR TASKS LIKE CLASSIFICATION, REGRESSION, ANOMALY DETECTION, PREDICTION, AND MORE

# Rapidminer



RAPIDMINER IS A FREE OPEN-SOURCE DATA SCIENCE PLATFORM THAT FEATURES HUNDREDS OF ALGORITHMS FOR DATA PREPARATION, MACHINE LEARNING, DEEP LEARNING, TEXT MINING, AND PREDICTIVE ANALYTICS.

# THANK YOU FOR WATCHING