



## Measuring Software Engineer Report

By

Harvey Gallagher

18322598



# **Table of Contents**

## **Introduction**

### **1.Measurable Data**

### **2.The different computational platforms used in this work**

### **3.Algorithmic approach**

### **4.Ethical concerns**

### **5.Conclusion**

## **Bibliography and References**

## **Introduction**

In this report, I will discuss the different ways in which software engineering can be measured and interpreted. The measure of Software Engineering is the measurement of a software product or a process software. This is an area of software engineering. Software Engineering can be measured in terms of measurable data and the computational tools used to perform the measurement. The different algorithm approaches that are used and the number of different ethics concerns surrounding the different type of analytics used to monitor software engineers. These are the various aspects of software engineering measurement. Software engineering is a subset of type of engineering, software engineers use the principles of engineering to test and construct applications to help users. In the last 20 years the idea of discipline has developed and grown hugely. This development has implemented the need of measuring accurately and interpreting these results. In this report, I will examine how discipline is measured in terms of

1. Measurable data
2. The different computational platforms needed for this work
3. The algorithm approaches
4. The ethics concerns regarding these analytics.

## **1.Measurable Data**

The area of the software engineering has developed and revolutionized over the last twenty years. The collecting of accurate a measurable data has become an integral part of understanding the work of each software engineer. In the early days the one of the biggest challenges was how and what data was going to be collected. This led the collection of data manually which was inaccurate and difficult to analyse. The Software developers were unsure of what to do with this data, this led to little data being collected.

The workplace began to realize the significance of monitoring work which in turn led to improvements in the workplace. The correlation between the measurement and improvement of is extremely helpful tool to measure the data for any profession. This data can provide companies with a better understanding of how their employees are performing. Companies can also minimise costs, time and maximise efficiency and productivity. The rapid growth of technology through communication, information and systems. By measuring the work of each software engineer, different areas of their work can be developed and improved based on the followings. Managerial decisions can be made in order to improve the performance of a company.

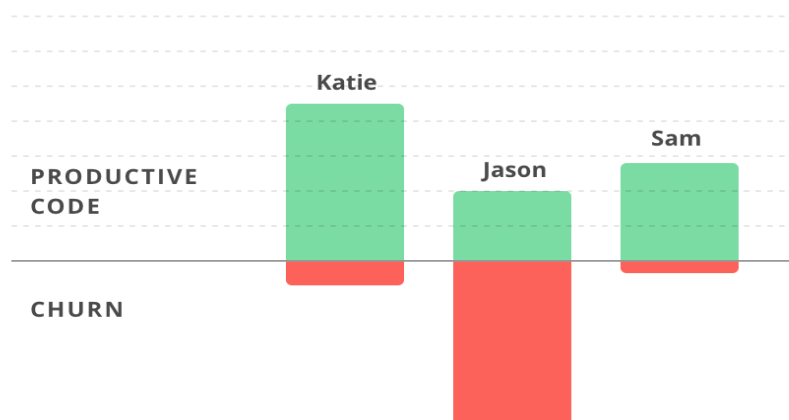
By measuring software there is no single way method which measures the production levels of a particular employee. The production level is determined on how developed the

software is , that be using to measure the data. Measuring processes are both extremely important and challenging for software engineers. The 3 categories of metrics used to measure software engineering processes: process metrics, quality metrics and productivity metrics. A software metric is where the information is where data from an application development lifecycle is measured and is used to examine the developers productivity. The data can originate from a single or multiple data source .

## Productivity Metrics:

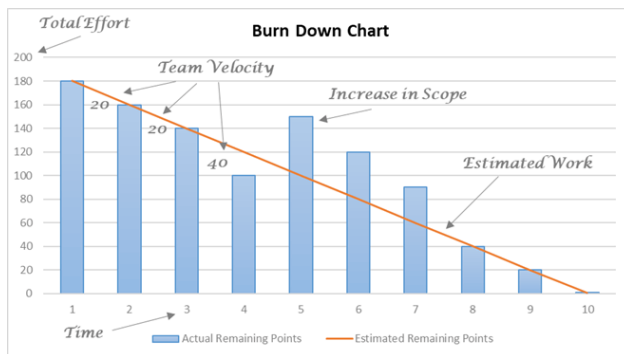
Software productivity is defined as the ratio between the functional values of the software that is produced to the effort and the expense required for the growth and development.

1.**Code churn** is the percentage of a developer's own code showing an edit or modify to their own work .To measure the modification we measure the total lines of code that were modified , added and deleted over a period of time. Code churn is regarded to be “non-productive work”, but high churn can be used in other aspects such as: indecisive stakeholders, unclear problems one is that are difficult to solve. Code Churn is best implemented for a developer when the team is made, and the churn moves from baseline that is set.



## 2.Sprint Burndown

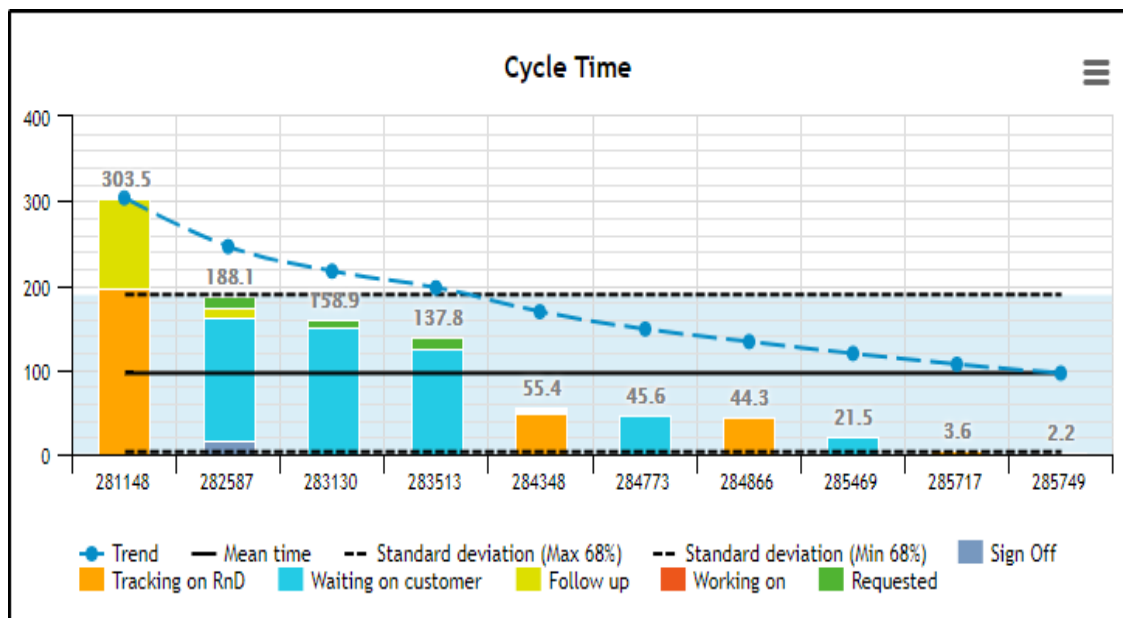
A project is not considered as several tasks and that all tasks must be carried out equally. Teams use the idea of story points. A story point is a high-level estimation of complexity in software. Sprint burndown illustrates perspective work based on the story points.



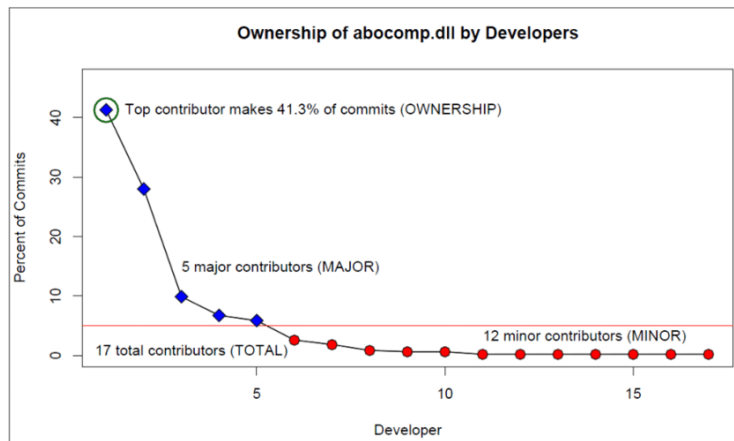
## Process Metrics

Process metrics illustrates the performance of the software engineers.

- Lead time** is the difference the initial and complete steps of the process. From a software engineering point of view, it is the quantity of the time that has passed between the identification of the task and the completion. Lead way of keeping track of how often the software developer is responding to the customer.



**2.The number of commits** is the measurement of the amount of time a software engineer spends on writing and modifying their own code. This way of measuring work activity is a good way but using this type of measuring for productivity is inaccurate .The larger the number of commits does not correlate with improvement in productivity and progress for the code to work.



## Quality Metrics:

**1.Test Coverage ratio** is calculated by getting the ratio between the total lines of code in a software and the number of lines used to test the code. A high test coverage program usually measured as a percentage ,has had more of it's source code executed during the testing , which means that it reduces the chance of containing undetected software bugs compared to a program with a lower test coverage.

**2.Application crash rate (ACR)** is calculated by dividing how many times an application by the amount of times the application was used. The number of ways to assess the way of measuring . The first way of assessing this metric is app crashes per user and an acceptable range for this metric can be less than 1%.The second way is the app crashes per session and an acceptable range is about less than 0.1%. The final way of assessing is the app crashes per screen per viewer which has range of less than 0.01%.

## Limitations :

Managers simply cannot read these metrics and come to conclusions in an instant. Metrics do not assume certain cases .These means that managers must take time to process and information and investigate the real story behind the numbers and what they represent.

We concluded that the existing models are incapable of predicting defects accurately using size and complexity of metrics. These models offer no coherent explanation of how defect introduction and detection affect defect count.(Fenton,N.E.,and Martin,N.1999)

## **2.The different computational platforms used for this work:**

The quality and efficiency of code is becoming hugely significant in the development of software, software organizations must be able to collect data in order to build useful software. Many analytic companies have come to the forefront to process this data. The job of these companies is to provide software that will monitor the code and will try to eliminate bugs and security leaks. Previously I have talked about measurements that these companies use, so here are some examples of these software analytics companies.

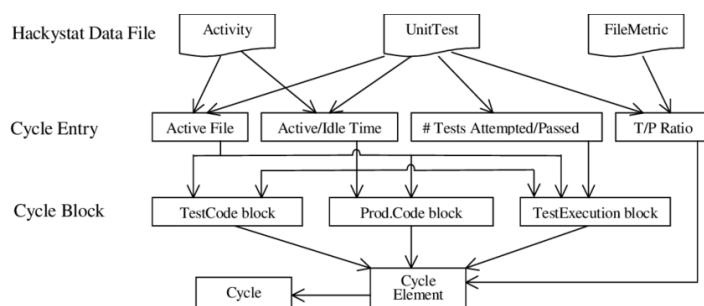
### **2.1Personal Software Process:**

PSP is a software development process that was created to help the improvement of software engineer's performance by using data disciplined. The PSP was setup by Watts Humphrey who taught through analysis of the application in a practical manner. Over the years there has been numerous changes to PSP. The engineers fill out a number of forms, summaries, time checklist. The collection and interpretation of the data is very time consuming and difficult as the data was gathered manually. This type of collection of data is also more prone to human error. This will lead to inaccurate results and poor-quality set of results. PSP however provides essential software engineering for data quality problems. Unfortunately PSP from a time invested perspective does provide a good return.

### **2.2Hackystat :**

Hackystat is an automated platform tool. Software developers use Hackystat in the form of sensors to the tools on the platform. These sensors gather the information and process it. The information is analysed, one of the features of Hackystat is that users shouldn't notice collection of their data and the ability to gather the data from customer and server side. The most appealing aspect is its fine grain data collection, this feature collects the data instantaneously, and tracks developers when they modify code. This is probably one of the main benefits of Hackystat however many developers find it uncomfortable of the amount of data that is collected.

According to (Johnson 2013). The nature of the collection allows managers to have a deeper understanding of the work that is being carried out. Many developers were uncomfortable with the accessibility platform and the platform gave others to work.



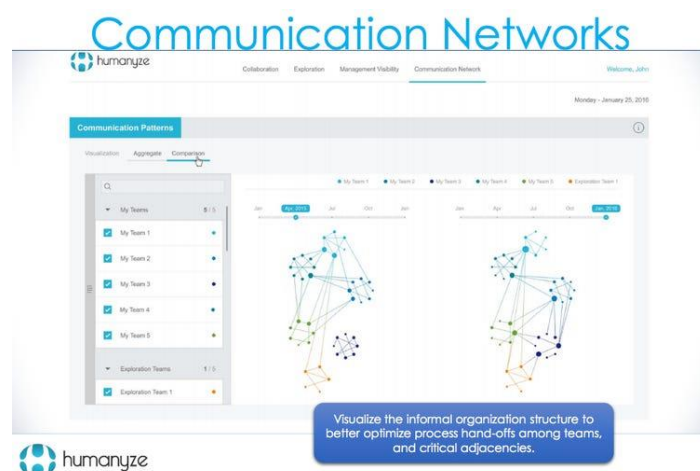
### 2.3 Code climate

In 2011, code climate was founded in order to solve problems. Two experienced software engineers Noah Davis and Bryan Helmkamp. This was founded as you move further away from the code to ensure software quality. Code climate is used by developers to make their code easier to interpret and makes sure the coders don't repeat the same code. The application also allows the code to be modified easily.



### 2.4 Humanyze

**Humanyze** provides companies software (sensors) to keep up to date of their employees' activities throughout the day. They have software such as Bluetooth to track location, accelerometer to keep track of movement and infrared to recognize which employee is speaking. Humanyze has become more prominent in the working world for example Bank of America, NASA and United States Army. Humanyze measures an organisation productivity instead of specific individuals. This may give an inaccurate picture of what the organisation represents.





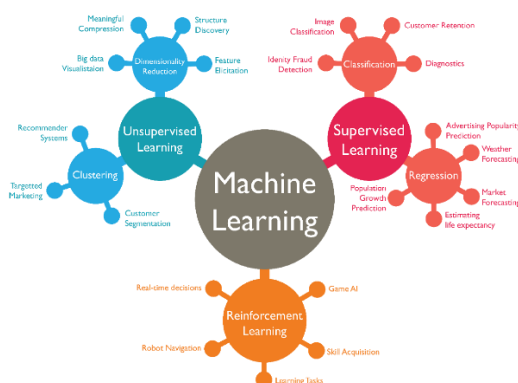
### **3.The Algorithm Approaches**

Data analysis is an instrumental part of software engineering. Algorithms has replaced the idea of manually analysis .The idea of algorithms are a lot more efficient and provide more information. Algorithms can provide information about patterns and trends. Companies use algorithms to analyse employee's performance, tech companies such Google and Facebook. Machine learning are integral part in using these algorithms. This allows computer software to interpret data and make decisions based on the information the software has been given. There are three types of ML algorithms

#### **1.Supervised Learning**

#### **2.Unsupervised Learning**

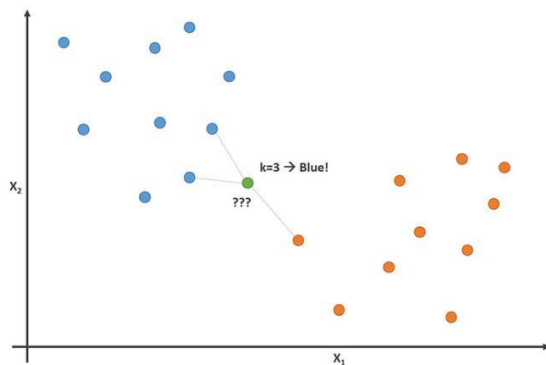
#### **3.Reinforcement Learning**



**3.1 Supervised Learning** is when an algorithm produces a function that has input that generates an output i.e.  $y=x$  where  $y$  is the output and  $x$  is the input. This is called supervised learning as data scientists gather results based on the outputs produced from the inputs put into the functions. The algorithm can then make predictions based on the results that are outputted. This process is repeated until the accuracy of the results are appealing to the data scientist. There are several learning algorithms such as k-nearest neighbours and linear discriminant analysis.

### **3.2 K – Nearest neighbours**

K – Nearest neighbours' algorithm is nonparametric function used for classification and regression. The algorithm works by the distance between points that represent pieces of data. The two points that are closest together are known as the k-nearest neighbours. The number of k-nearest neighbours for whatever value k is equal to. For example, if k=2 the software in the machine will choose the two closest observations to the point in the data that is referring to.



### **3.3 Linear Discriminant Analysis**

Linear Discriminant Analysis is a method used in statistics in order to find a linear combination that characterizes or separates two or more classes. The algorithm estimates the probability of inputs and produces outputs. Data scientists can learn and interpret the results based on the outputs produced.

### **3.4 Unsupervised Learning**

Unsupervised learning is very similar to machine learning, as the computer's software aim is to understand how to do something without telling the user. The algorithm's aim is to process information on its own without the guidance of the human brain. The input is placed into the function and the output is unknown. That is the difference in unsupervised learning. Two examples of unsupervised learning are k-means clustering and Principal component analysis.

### **3.5 K-Means Clustering**

K-mean clustering are very useful in data analytical world, can be divided into a set of clusters depending on its attributes and characteristics. The algorithm divides two different categories in the one group like for example a group of fast sprinters and slow sprinters. The data are divided into k groups so the observations with similar characteristics are in the same group.

### **3.6 Principal Component Analysis**

Principal Component analysis is the process of inputting the principal components and use them to perform changes in the data. CA is used in exploratory analysis and making predictions based on the results of the data. The aim is to identify the variables that represent the variance in the data. This can represent correlation in the data.

### **3.7 Reinforcement Learning**

Reinforcement Learning is a type of learning where the algorithm makes a sequence of decision based on the interaction with the environment. This algorithm is hugely based on decision making. This type of ML is becoming hugely important in the world of technology, especially in driverless cars where decision making is the key to success. The idea of the algorithm is to learn from previous mistakes and not to make the same ones in the future.

The only limitation to this algorithm is that as the inputs are put into the algorithm the results get more accurate, however the results could improve so much that the human mind cannot interpret results and will lead to human error. The development of ML has given rise to a range of new opportunities.

## **4. The ethics concerned about these type of analytics**

The measurement and assessment of software engineering has developed and grown hugely through using data analytics and machine learning. The big question remains is it all ethical? The idea of right and wrong remains a big question in the world of software engineering. The amount of data being collected is it ethical? The algorithms mentioned previously all require data, but require data where privacy is sometimes breached. For example, the whole Cambridge Analytica Scandal. Are firms collecting too much information?

### **4.1 Data Collection**

Data collection is one of the main aspects of software engineering that raises many questions about ethics. Developers can sometimes feel uncomfortable about the collection of data that they do not know that is being collected. Humanzyne technology as discussed before can track their employees every movement and interaction with each other. This can affect the employees morale and confidence in their own work. If I was in that situation personally I would feel stressed and would prevent me from reaching my full potential of work level. Companies need to learn how to manage to balance between gathering enough information that doesn't breach privacy of the employees. My feelings are very similar to the

idea of Somerville's(2011) believes that honesty and integrity is the key to helping software engineers perform to the best to their ability.

#### **4.2 Data Usage**

The large amount of data that is collected by companies about their employees. Companies can do analytics on each of their employees which is called predictive people analytics , this tool predicts what employees are going to perform well and which are most likely to leave. This kind of leaves the question, is this fair? Does this give every employee a fair chance to prove they are worthy of the job? This also leaves a debate are the algorithms a 100% accurate as it may give employers false information.

#### **4.3 Data Sovereignty**

Data sovereignty is the idea that information is stored after being converted into binary form .The problem is that many countries have different privacy regulations. So, information that is stored in another country cannot be used in court of another country if an employee presses charges as regards privacy. Employers have complete trust in employees not to share information and should adhere to the privacy regulations.

### **5. Conclusion**

In conclusion ,I have outlined the various ways that software is measured on performance and how that the data in analysed and interpreted. From writing this report I have clearly learned that there are huge benefits from measuring the performance of a software engineer. This allows head of companies to make vital decisions from the productivity and quality of the results. The variety of software engineer metrics used to measure performance can be hugely significant in adding value to an organisation. As long these metrics are used in the correct manner as regards measuring in an accurate precise way. According to (Hackernoon 2018) it is easy to focus on the metrics . However, to be a successful software engineer one needs several attributes such as patience, open mindedness , good communication, good time management and creativity. These attributes should all be assessed when measuring the performance of the software engineer.

## **Bibliography and References**

[https://eecs.qmul.ac.uk/~norman/papers/jss\\_1999.pdf](https://eecs.qmul.ac.uk/~norman/papers/jss_1999.pdf)

[https://learn.pluralsight.com/resource/offers/2020/data-driven-engineering?oid=7010c0000022ShUAAU&utm\\_term=&aid=7014Q0000022dwUQAQ&promo=&utm\\_source=non\\_branded&utm\\_medium=digital\\_paid\\_search\\_google&utm\\_campaign=B2B\\_NA\\_Content\\_Data-Driven-Engineering\\_BMM&utm\\_content=&gclid=CjwKCAiA5IL-BRAzEiwA0lcWYjWUEVG1N0dXPI7IV9RdjgH8LKcaDpgyrddM15EqDZDQv9J5UTAuAxoCxi4QAvD\\_BwE](https://learn.pluralsight.com/resource/offers/2020/data-driven-engineering?oid=7010c0000022ShUAAU&utm_term=&aid=7014Q0000022dwUQAQ&promo=&utm_source=non_branded&utm_medium=digital_paid_search_google&utm_campaign=B2B_NA_Content_Data-Driven-Engineering_BMM&utm_content=&gclid=CjwKCAiA5IL-BRAzEiwA0lcWYjWUEVG1N0dXPI7IV9RdjgH8LKcaDpgyrddM15EqDZDQv9J5UTAuAxoCxi4QAvD_BwE)

[https://hyperscience.com/document-processing-automation/?gclid=CjwKCAiA5IL-BRAzEiwA0lcWYjWUEVG1N0dXPI7IV9RdjgH8LKcaDpgyrddM15EqDZDQv9J5UTAuAxoCxi4QAvD\\_BwE#evaluation?utm\\_campaign=global-2020q3-assessing-goals-campaign&utm\\_medium=paidsearch&utm\\_source=adwords&utm\\_content=ungated-guide-document-processing-automation&bt=479269656473&bk=process%20automation%20software&bm=b&bn=g&bg=110983508857](https://hyperscience.com/document-processing-automation/?gclid=CjwKCAiA5IL-BRAzEiwA0lcWYjWUEVG1N0dXPI7IV9RdjgH8LKcaDpgyrddM15EqDZDQv9J5UTAuAxoCxi4QAvD_BwE#evaluation?utm_campaign=global-2020q3-assessing-goals-campaign&utm_medium=paidsearch&utm_source=adwords&utm_content=ungated-guide-document-processing-automation&bt=479269656473&bk=process%20automation%20software&bm=b&bn=g&bg=110983508857)

[https://www.google.com/search?q=k+nearest+neighbor&tbm=isch&ved=2ahUKEwjYwOGB-KLTAhX5VhUIHcHKAfcQ2-cCegQIABAA&oq=k+nearest+&gs\\_lcp=CgNpbWcQARgAMgIIADICCAAYAggAMgIIADICCAAYAggAMgIIADICCAAYAggAMgIIADoECCMQJzoHCCMQ6gIQJzoFCAAQsQM6CAgAELEDEIMBOgQIABBDUlzECFim5whgjIlaAFwAHgDgAH1AogBIRKSAQgxNy4zLjEuMZgBAKABAaoBC2d3cy13aXoWtaW1nsAEKwAEB&sclient=img&ei=fw3BX5iFJfmt1fAPwcmHuA8&bih=647&biw=686&rlz=1C1CHBF\\_enIE815IE815#imgsrc=hG\\_sPMaHwgZt0M](https://www.google.com/search?q=k+nearest+neighbor&tbm=isch&ved=2ahUKEwjYwOGB-KLTAhX5VhUIHcHKAfcQ2-cCegQIABAA&oq=k+nearest+&gs_lcp=CgNpbWcQARgAMgIIADICCAAYAggAMgIIADICCAAYAggAMgIIADICCAAYAggAMgIIADoECCMQJzoHCCMQ6gIQJzoFCAAQsQM6CAgAELEDEIMBOgQIABBDUlzECFim5whgjIlaAFwAHgDgAH1AogBIRKSAQgxNy4zLjEuMZgBAKABAaoBC2d3cy13aXoWtaW1nsAEKwAEB&sclient=img&ei=fw3BX5iFJfmt1fAPwcmHuA8&bih=647&biw=686&rlz=1C1CHBF_enIE815IE815#imgsrc=hG_sPMaHwgZt0M)

[https://www.google.com/search?q=companies+who+use+humanyze&rlz=1C1CHBF\\_enIE815IE815&oq=companies+who+use++humanyze&aqs=chrome..69i57j33i160.13305j0j7&sourceid=chrome&ie=UTF-8](https://www.google.com/search?q=companies+who+use+humanyze&rlz=1C1CHBF_enIE815IE815&oq=companies+who+use++humanyze&aqs=chrome..69i57j33i160.13305j0j7&sourceid=chrome&ie=UTF-8)

[https://www.google.com/search?q=machine+learning+algorithms&rlz=1C1CHBF\\_enIE815IE815&sxsrf=ALeKk00LI-VIIiSqYJuyIndnF2inH7l8lw:1606420421842&source=lnms&tbm=isch&sa=X&ved=2ahUKEwiT8em0\\_qDtAhVrURUIHbvXCmAQ\\_AUoAXoECAQQAw#imgsrc=1usR\\_RY9LI\\_8pM](https://www.google.com/search?q=machine+learning+algorithms&rlz=1C1CHBF_enIE815IE815&sxsrf=ALeKk00LI-VIIiSqYJuyIndnF2inH7l8lw:1606420421842&source=lnms&tbm=isch&sa=X&ved=2ahUKEwiT8em0_qDtAhVrURUIHbvXCmAQ_AUoAXoECAQQAw#imgsrc=1usR_RY9LI_8pM)

[https://www.google.com/search?rlz=1C1CHBF\\_enIE815IE815&sxsrf=ALeKk02q0EMfInNXQepq4DI6vReqXp409A%3A1606423064794&ei=GBLAX5GEMO-X1fAPn-CF0Aw&q=linear+discriminant+analysis&oq=linear+discriminat+anay&gs\\_lcp=CgZwc3ktYWlQARgAMgIsIABDJAXAwEAoQHjoHCAAQRxCwAzoFCAAQyQM6BggAEByQHjoHCCMQ6gIQJzoECCMQJzoECC4QJzoICAAQyQMqQkQI6BQgAEJECoggIABCAxCDAToICC4QsQMqgwE6DgguELEDEIMBEMcBEKMCogQIABBDogQILhBDOgoIABCAxCDARBDogclABDJAXBDogclABCAxABDOgIIADoNCAAQsQMqgwEQyQMqQzoICC4QxwEQrweE6BwgAEMkDEAo6BAGAEApQu-](https://www.google.com/search?rlz=1C1CHBF_enIE815IE815&sxsrf=ALeKk02q0EMfInNXQepq4DI6vReqXp409A%3A1606423064794&ei=GBLAX5GEMO-X1fAPn-CF0Aw&q=linear+discriminant+analysis&oq=linear+discriminat+anay&gs_lcp=CgZwc3ktYWlQARgAMgIsIABDJAXAwEAoQHjoHCAAQRxCwAzoFCAAQyQM6BggAEByQHjoHCCMQ6gIQJzoECCMQJzoECC4QJzoICAAQyQMqQkQI6BQgAEJECoggIABCAxCDAToICC4QsQMqgwE6DgguELEDEIMBEMcBEKMCogQIABBDogQILhBDOgoIABCAxCDARBDogclABDJAXBDogclABCAxABDOgIIADoNCAAQsQMqgwEQyQMqQzoICC4QxwEQrweE6BwgAEMkDEAo6BAGAEApQu-)

[dOWOqiT2D9tU9oAnAAeASAAcwCiAGplpBCTEwLjEwLjcuMZgBAKABAaoBB2d3cy13aXqWAQrIAQjAAQE&sclient=psy-ab](https://doi.org/10.1007/978-3-642-30287-9_20)

[https://link.springer.com/chapter/10.1007/978-3-642-30287-9\\_20](https://link.springer.com/chapter/10.1007/978-3-642-30287-9_20)

<https://www.computer.org/csdl/magazine/so/2013/04/mso2013040057/13rRUxC0SCb>

[https://en.wikipedia.org/wiki/Software\\_engineering](https://en.wikipedia.org/wiki/Software_engineering)

[https://en.wikipedia.org/wiki/Principal\\_component\\_analysis](https://en.wikipedia.org/wiki/Principal_component_analysis)

[https://en.wikipedia.org/wiki/K-nearest\\_neighbors\\_algorithm](https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm)

[https://en.wikipedia.org/wiki/K-means\\_clustering#:~:text=k%2Dmeans%20clustering%20is%20a,a%20prototype%20of%20the%20cluster.](https://en.wikipedia.org/wiki/K-means_clustering#:~:text=k%2Dmeans%20clustering%20is%20a,a%20prototype%20of%20the%20cluster.)

[https://en.wikipedia.org/wiki/K-means\\_clustering#:~:text=k%2Dmeans%20clustering%20is%20a,a%20prototype%20of%20the%20cluster.](https://en.wikipedia.org/wiki/K-means_clustering#:~:text=k%2Dmeans%20clustering%20is%20a,a%20prototype%20of%20the%20cluster.)