

---

**Program Name:** Big Data Analytics

**Project Code:** CPL-5559-DSMM

---

## AIP Project Case Study Orientation

### Vocational Learning Outcomes (VLOs) Covered in this WIL Project Case Study

1. Collect, house, extract, manipulate, maintain, and mine data sets that respond to organizational, financial, or market needs.
2. Recommend different systems and network architectures, artificial intelligence, and data storage technologies to support data analytics and Big Data.
3. Design and apply data models that meet the needs of a specific operational process or business model.
4. Develop software applications, algorithms, and artificial intelligence models to manipulate, correlate and reduce data sets and produce project documentation and reports.
5. Collaborate effectively with diverse teams to design and present data visualizations that communicate Big Data concepts and information to stakeholders and business professionals.
6. Apply business analytics, business intelligence tools and research to support evidence-based decision-making that helps an organization meet their stated objectives.
7. Identify and assess data analytics and Big Data business strategies and workflows to respond to new opportunities or provide project solutions.
8. Implement data security solutions in compliance with corporate security policies, ethical standards, and industry regulations.
9. Deliver data-oriented projects using data science, business analysis, and project management principles, tools, and techniques.
10. Develop artificial intelligence solutions to support administration, decision-making, planning, risk management, logistics, manufacturing, smart devices, and robotics.

---

**Program Name:** Big Data Analytics

**Project Code:** CPL-5559-DSMM

---

## Essential Employability Skills (EESs) Covered in this AIP Project Case Study

- Communication
  - It helps to communicate clearly, correctly, and concisely in different forms. These include oral, written, and visual.
- Numeracy
  - This skill set helps to solve mathematical operations effectively with accurate precision.
- Critical thinking & problem solving
  - It is a systemic approach to attempting to resolve problems by analyzing the pros and cons of a decision.
- Information management
  - It helps to locate, select, organize, and document information with the use of technology by analyzing aspects and gathering information from a variety of sources.
- Interpersonal Skills
  - This skill set is important as it helps to respect others' opinions or input. It helps to build teams and maintain relationships to achieve overall team or organizational goals.
- Personal Skills
  - These soft skills are important in developing employability talents, such as dependability, adaptability, and problem-solving skills.

**Program Name:** Big Data Analytics

**Project Code:** CPL-5559-DSMM

---

## Week 1

### This Week's Detailed Case study Information

Congratulations you have been hired as an ML Engineer for Mountain Analytics located in Toronto ON.

Mountain Analytics is an analytics company that provides professional services including but not limited to data collection, data organization in various domains such as sales forecasting, social media analysis and digital marketing. In addition, Mountain Analytics performs a prescriptive and predictive analysis.

Fox studio is a company which sponsors movies and is trying to analyze the reviews related to movies. The data is collected from various sources such as websites and forums. People are writing about the plot, visual effects, characters and many more on these sites. The data from these sources are collected and provided by Fox studio as per the contract between Mountain Analytics and Fox studio and the objective of this project is to analyze the sentiments in the given data. As per the contract, the name of the movies provided will not be revealed. The data is stored in on-premise infrastructure as well as the cloud. Dynamodb is used to store data in the cloud.

Fox studio is trying to figure out the public's sentiments toward movies in order to increase revenue from the movies. The goal behind this project is to find out which genre of movies are getting popular and what are the sentiments for these movies. Based on the findings of the movies, fox studio will invest in certain kinds of movies in future. To facilitate informed decision making this project is started, and there will be other factors affecting the final decision for investment. Your team is responsible for finding the sentiment in the review data. Since the data is present on both on-premise and cloud your team's responsibility is to find the best ways to collect the data and perform sentiment analysis in it.

Your Analytics team consists of

- Josh Mosby (Lead Data Scientist)
- Marshal Segel (Data Scientist)
- Barney Haris (Data manager)
- Robin Smulders (Project Manager)
- Rahul Nayar (Data Engineer)

You will be working closely as an ML Engineer with these team members.

### Week 1 Onboarding Expectations and Participation

Your task this week is to participate in training and orientation for Mountain Analytics. You will participate in a variety of exercises that are designed to get to know you better and

**Program Name:** Big Data Analytics

**Project Code:** CPL-5559-DSMM

---

understand your role within the team. You will participate in Team-building exercises that prepare you for success within the Project. As with any position, you will have an excellent opportunity to build on your skills as a leader so long as you put forth your best effort. Use this week to develop a communication plan with your team and be ready to dive into the deliverables starting next week.

**Note:** You can make any assumptions that are deemed necessary for each case on a week-by-week basis. You will not be provided direct answers or 100% of the information necessary to complete each deliverable. Instead, focus on delivering the highest quality outcome possible to highlight your talent as a group. You would be presenting these deliverables to Robin and would want to ensure that the work is of the highest quality.

This section will be available to you for the entirety of the project. However, each subsequent week's case study information may only be available for that week. Be sure to download and save this week's information for future use.

**Program Name:** Big Data Analytics

**Project Code:** CPL-5559-DSMM

---

## Week 2

### Applicable VLOs or EEs for This Week's Case Study

- Collect, house, extract, manipulate, maintain and mine data sets that respond to organizational, financial, or market needs.
- Recommend different systems and network architectures, artificial intelligence, and data storage technologies to support data analytics and Big Data.

### This Week's Detailed Case Study Information

Before diving into the project, the input, Tools & Technologies, and output (ITTO) should be identified. For this project, ITTO will be as follows

- Input: Datasets provided by Fox Studio
- Tools & Technologies: **Up to your team.**
- Output: Prediction of sentiments.

Different sets of tools and technologies can be used to solve the same problem. Before deciding on the technology stack consider the skillset of the team and the ideal stack to solve the problem. You need to find the balance between these factors to solve the problems. As an ML Engineer, you should be up to date with emerging technologies their alternatives and their limitations

One of the requirements for any kind of data solution is domain knowledge. Knowing in and outs of the data lifecycle for a specific application will help you down the road. So, it's better to gather knowledge about the domain of the data.

- Research about the tools services and software components required to solve the problems.
- Hadoop and Its Ecosystems such as Spark, Scala, Kafka, HBase, AI

### Deliverables for This Week's Case Study

1. Analyze how ratings and reviews are collected in IMDB and rotten Tomatoes as a part of domain knowledge.
2. Prepare a brief report highlighting the jargon used in the movies.
3. Create a block diagram of the IT infrastructure required for this system.

---

**Program Name:** Big Data Analytics**Project Code:** CPL-5559-DSMM

---

## Week 3

### Applicable VLOs or EEs for This Week's Case Study

- Recommend different systems and network architectures, artificial intelligence, and data storage technologies to support data analytics and Big Data.
- Design and apply data models that meet the needs of a specific operational process or business model.
- Develop software applications, algorithms, and artificial intelligence models to manipulate, correlate and reduce data sets and produce project documentation and reports.

### This Week's Detailed Case Study Information

It has been found that a project in which a significant amount of time was spent in the planning phase has a higher chance of success. In the planning phase, the scope, cost and time of the project are finalized and the Work BreakDown structure document is created. To finalize this trio scope, cost and schedule you will be attending a meeting with Josh, Robin and Barney. In most cases, data are gathered from various sources, as usual for this project as well data are stored in different locations on-premises and cloud. Mountain Analytics is trying to migrate everything from its on-premises infrastructure to the cloud and Josh believes that it is the right time to do so. Data migration is the process of moving data from one location to another, from one format to another, or from one application to another. Generally, this is the result of introducing a new system or location for the data. The business driver is usually an application migration or consolidation in which legacy systems are replaced or augmented by new applications that will share the same dataset. These days, data migrations are often started as firms move from on-premises infrastructure and applications to cloud-based storage and applications to optimize or transform the company.

### Deliverables for This Week's Case Study

1. Try out commands used for reading and writing data to HDFS.
2. Create an AWS account.
3. Identify the different file formats to be used for this use case
4. Get a sample data file from HDFS to your local machine
5. Write a python script to convert the parquet file to CSV and vice versa and test it.
6. Write business objectives with which Mountain Analytics can achieve specified goals, increase productivity and decrease human labour. Write ways to increase the efficiency of the project. Apart from that how proposed technological tools can be used from a security point of view?

---

**Program Name:** Big Data Analytics**Project Code:** CPL-5559-DSMM

---

## Week 4

### Applicable VLOs or EEs for This Week's Case Study

- Develop software applications, algorithms, and artificial intelligence models to manipulate, correlate and reduce data sets and produce project documentation and reports.

### This Week's Detailed Case Study Information

Robin suggested that before you actually start the project you need to know about the overall work to be done in the project. It helps to manage the resources and time for the various task and track the progress. The identification of task helps to divide the work into smaller chunks which are more tangible and manageable. The smaller chunk of work is called the work package and can be achieved with the help of the document called Work Breakdown Structure (WBS).

Firstly, you will assist Robin to identify the limitations and risks that need to be addressed to ensure the project's ultimate success. The three primary constraints that project managers should be familiar with are time, scope, and cost. These are frequently known as the triple constraints or the project management triangle. Each constraint is connected to the other two; so, for example, increasing the scope of the project will likely require more time and money, while speeding up the timeline for the project may cut costs, but also diminish the scope.

Furthermore, you will create a work breakdown Structure document to divide the work into smaller chunks. The work package will be assigned to the team or individual to achieve the goal. Your team is planning to work on an Agile environment so that the work is more transparent and the changes can be accommodated as you move forward on the project.

Robin has requested your work on the identification of these constraints for this project. You can always consult with experts when in doubt. There might be different expertise for each of these constraints.

### Deliverables for This Week's Case Study

1. Identify every kind of resource needed for this project (Human, computing, cost)
2. Review the WBS of a similar project and create a Work Breakdown Structure(WBS) document including details of the task to be done on every phase of this project.
3. Estimate the time to complete every work package in the WBS document and create a milestone plan with expected deliverables.

---

**Program Name:** Big Data Analytics

**Project Code:** CPL-5559-DSMM

---

## Week 5

### Applicable VLOs or EEs for This Week's Case Study

- Develop software applications, algorithms, and artificial intelligence models to manipulate, correlate and reduce data sets and produce project documentation and reports.

### This Week's Detailed Case Study Information

Cloud computing is the delivery of computing services—including servers, storage, databases, networking, software, analytics, and intelligence—over the Internet ("the cloud") to offer faster innovation, flexible resources, and economies of scale. You typically pay only for cloud services you use, helping you lower your operating costs, run your infrastructure more efficiently, and scale as your business needs change.

To solve a typical problem, there are multiple tools and services in the cloud which can be utilized. Any specific tools won't be a perfect match for every use case scenario. You must identify the needs of your project, build a resource plan and decide on the services to use. Robin has asked you to find the services to be used for this project. At this phase of the project, you will need a storage service and a database. Make sure whichever service you choose, you should be able to explain the reason behind your choice.

**Hint:** DyanmoDb is used as a Database, so choosing some service within the AWS domain would be a wise decision. Refer to the documentation of the cloud services in case of confusion. To create a table in the DynamoDB you need to know the structure of the data. The data may be zipped or raw. There are different possible file formats and before diving into ETL processes make sure you are clear about the file formats and structure.

- Explore the format of your data and identify the required field
- Analyze the data type required for each field
- Research the use case scenario of using DynamoDB and other NoSQL databases.

### Deliverables for This Week's Case Study

1. Using the account you created in earlier week, create S3 and DynamoDB
2. Upload the data to AWS S3 and export it to DynamoDB
3. Use the boto library in python to connect with DynamoDB and test the connection
4. Display the first ten rows in DynamoDB with python API.



**Program Name:** Big Data Analytics

**Project Code:** CPL-5559-DSMM

---

## Week 6-Midterm Week

### Mid-Term Panel Evaluation Preparation

The team will prepare and present the Mid-Term Panel Evaluation this week. For the Team Presentation create a professional multimedia presentation highlighting the key aspects of your project thus far. Please see Moodle for full details.

#### **Presentation**

##### CONTENT

- Overview of work in the Mountain Analytics
- Highlight three key areas you find of interest:
  - Two areas related to weekly work completed
  - One area to highlight PD or other activity
- Apply reflecting skills
- Present the importance/Benefit of work to Mountain Analytics.

##### FORMAT & LAYOUT

- Research and design professional presentations with a popular tool such as PowerPoint, Canva or Adobe, free of spelling and grammatical errors. Images and other graphics used should represent the company and communicate the brand's color scheme.
- Team effort: every team member needs to present, contribute and account for the contribution towards the presentation

---

**Program Name:** Big Data Analytics**Project Code:** CPL-5559-DSMM

---

## Week 7

### Applicable VLOs or EEs for This Week's Case Study

- Implement data security solutions in compliance with corporate security policies, ethical standards, and industry regulations.
- Identify and assess data analytics and Big Data business strategies and workflows to respond to new opportunities or provide project solutions.

### This Week's Detailed Case Study Information

This week you will be collaborating with Barney to maintain data security. Barney is well aware of different kinds of cyber-attacks and breaches. He tries to keep everything updated and secure and put lots of effort into doing so. There are many reasons why data security is important to organizations in all industries all over the world. Organizations are legally obliged to protect customer and user data from being lost or stolen and ending up in the wrong hand. A data breach is ever existing risk for data analysis projects. Multiple companies have ended up paying millions of dollars in fines because of data breaches. So Barney tries to avoid such a case at any cost.

Risk planning is the process of identifying, prioritizing, and managing risk. Every project or initiative has objectives, that is, goals that it seeks to accomplish. These are often called Critical Success Factors (CSF). Risk events threaten the successful completion of these critical success factors. Thus, risk planning involves identifying the most important risk events in advance, prioritizing them, and developing the appropriate risk response plans. There are three steps to risk planning: Identifying, prioritizing, and determining the response. For this project, data security is a major risk, and your team must try to avoid data breaches and other involved risks.

### Deliverables for This Week's Case Study

1. List the approaches to maintain the security of the data if the infrastructure is based on-premises.
2. Identify the roles and permissions required to do these tasks in the cloud services?
3. Will built-in roles be enough to fulfil your requirement, or do you need a custom role to access the services?
4. List the approaches to maintain the security of the data if the infrastructure is based on-premises.
5. Propose security tools to prevent any security-related issues in the cloud.
6. Write a short report on how PII can be tied to security threats.

---

**Program Name:** Big Data Analytics**Project Code:** CPL-5559-DSMM

---

## Week 8

### Applicable VLOs or EEs for This Week's Case Study

- Develop software applications, algorithms, and artificial intelligence models to manipulate, correlate and reduce data sets and produce project documentation and reports.
- Develop artificial intelligence solutions to support administration, decision-making, planning, risk management, logistics, manufacturing, smart devices, and robotics.

### This Week's Detailed Case Study Information

Supervised learning, also known as supervised machine learning, is a subcategory of machine learning and artificial intelligence. It is defined by its use of labelled datasets to train algorithms to classify data or predict outcomes accurately. As input data is fed into the model, it adjusts its weights until the model has been fitted appropriately, which occurs as part of the cross-validation process. Supervised learning helps organizations solve a variety of real-world problems at scale. The dataset provided to you has labels on it. Now, you can initiate the model-building process, the built model will analyze the sentiments into negative and positive. Josh is skeptical about your skill sets for building models for AI/ML workflows. Prove him wrong by performing all the steps of the machine learning workflow and building the model that will identify the sentiments of reviews. Use a confusion matrix for evaluating the performance of the model and discuss the result with him.

Robin has asked you to calculate the cost of running the operation in the cloud and on-premise machine and present the report comparing the cost. Explore any cloud solutions that are available for this use case and make any suggestions for optimal solutions. It's always wise to use graphs and charts to compare budgets and resources in reports. Josh has asked you to Implement AWS S3, and DynamoDB for data collection purposes and Sage maker, notebooks and python scripts for analytical purposes.

### Deliverables for This Week's Case Study

1. Prepare an architecture diagram for the components used when implementing AWS.
2. Identify the different libraries of python to be used for data cleaning, transformation, and modelling processes.
3. Use a confusion matrix to evaluate your model performance and explain its significance in this use case.
4. Build a model using an ideal algorithm for this scenario and rely on the F1 score and accuracy to choose the model.

---

**Program Name:** Big Data Analytics**Project Code:** CPL-5559-DSMM

---

## Week 9

### Applicable VLOs or EEs for This Week's Case Study

- Design and apply data models that meet the needs of a specific operational process or business model.
- Collaborate effectively with diverse teams to design and present data visualizations that communicate Big Data concepts and information to stakeholders and business professionals.

### This Week's Detailed Case Study Information

Exploratory Data Analysis is one of the major steps to fine-tune the given data set(s) in a different form of analysis to understand the insights of the key characteristics of various entities of the data set like column(s), row(s) by applying Pandas, NumPy, Statistical Methods, and Data visualization packages. EDA will result in

- Understanding the given dataset and helping clean up the given dataset.
- It gives you a clear picture of the features and the relationships between them.
- Providing guidelines for essential variables and leaving behind/removing non-essential variables.
- Handling Missing values or human error.
- Identifying outliers.

Mountain Analytics has a Tableau subscription for its data visualization needs. This week you are collaborating with Marshal to perform EDA on the cleaned data. In earlier weeks you cleaned the data set, now it's time to explore more about the data set and find anomalies and outliers. You will be using both python libraries and Tableau to visualize the data and learn more about it. This step is a continuation of a model-building process that you started in earlier weeks. Data visualization helps to analyze data, and find hidden patterns, anomalies and outliers.

- Use different plots and graphs to visualize the data
- Can python libraries be used instead of Tableau in every use case?
- Build the ML model again with the insight you received after EDA with visualization.

### Deliverables for This Week's Case Study

1. Perform Exploratory Data Analysis (EDA) on the data and build the model using a classification algorithm.
2. List the insights about data you were able to see due to data visualizations.
3. Build a model after doing EDA with visualizations and compare the models.

---

**Program Name:** Big Data Analytics**Project Code:** CPL-5559-DSMM

---

## Week 10

### Applicable VLOs or EEs for This Week's Case Study

- Develop software applications, algorithms, and artificial intelligence models to manipulate, correlate and reduce data sets and produce project documentation and reports.
- Deliver data-oriented projects using data science, business analysis, and project management principles, tools, and techniques.

### This Week's Detailed Case Study Information

Data science is an iterative process. You have to go back and forth between components of the project. For example, your model performance shows promising results but you realized that a different feature engineering technique may provide even better accuracy.

You have developed a model using an algorithm to classify the transactions in weeks 8 and 9. Use different algorithms and compare the performance of models. The use of an algorithm depends on the use case, not all algorithms will be applicable to all classification problems. You will be discussing the choice of your algorithm with Josh and Marshal so you should get facts and talking points ready.

Mountain analytics is planning to integrate the developed model into a web application using REST API. At the end of every day, the system would send the batch data of review to the application and identify sentiments. The aim of the system is to track the sentiments of the customer on daily basis. Robin has asked you to come up with the IT infrastructure to make this solution possible.

- Research the cross-fold validation and evaluate your model using the same.
- Use different folds for k-cross-fold validation and analyze the result.
- Research how the built models can be saved and reused.

### Deliverables for This Week's Case Study

1. Prepare a brief report on API and types of API and architecture.
2. Build the block diagram for the application in which the model built can be integrated with the web application.
3. Make a report highlighting the performance of different models and explaining how cross-validation assists in measuring algorithm performance.
4. Research the ways to implement these models in real-time reviews and predict the possible outcomes.

---

**Program Name:** Big Data Analytics**Project Code:** CPL-5559-DSMM

---

## Week 11

### Applicable VLOs or EEs for This Week's Case Study

- Develop software applications, algorithms, and artificial intelligence models to manipulate, correlate and reduce data sets and produce project documentation and reports.
- Deliver data-oriented projects using data science, business analysis, and project management principles, tools, and techniques.

### This Week's Detailed Case Study Information

A machine learning pipeline is an end-to-end construct that orchestrates the flow of data into, and output from, a machine learning model (or set of multiple models). It includes raw data input, features, outputs, the machine learning model and model parameters, and prediction outputs. The design and implementation of a machine learning pipeline is at the core of enterprise AI software applications and fundamentally determines performance and effectiveness. In addition to the software design, additional factors must be considered, including the choice of machine learning libraries and runtime environments (processor requirements, memory, and storage).

In the case of normal machine learning workflow, the model is the product and in automated workflows, the pipeline is the product. Even an ad hoc model can be deployed in real time depending on the use case. Save the ad hoc model (without pipeline) using pickle and Joblib libraries. In earlier weeks you have tried to build an ad hoc model and the code might not be reusable it can be mostly used for research purposes and test environments. Building with the same fundamentals Marshal has asked you to build a pipeline for the same workflow, this time making the reusable code. Refer to Sklearn documentation to make pipelines.

To make the work reusable and implement the modular design in ML workflow create pipelines.

- Research about the ML pipelines and the Benefits of using pipelines
- Create functions for every step that you build in ML workflow in earlier weeks
- Create a pipeline for the whole process

### Deliverables for This Week's Case Study

1. Submit both the pickle and joblib file.
2. Submit the python scripts created for the pipeline implementation.

---

**Program Name:** Big Data Analytics

**Project Code:** CPL-5559-DSMM

---

## Week 12

### Applicable VLOs or EEs for This Week's Case Study

- Identify and assess data analytics and Big Data business strategies and workflows to respond to new opportunities or provide project solutions.

### This Week's Detailed Case Study Information

A pre-trained model refers to a model or a saved network created by someone else and trained on a large dataset to solve a similar problem. AI teams can use a pre-trained model as a starting point, instead of building a model from scratch. Training the model from scratch usually involves one or more of these components: making use of a widely accepted sentiment lexicon, scoring sentiment by human experts, labelling data by agency contractors or research assistants, and tuning the model that performs well on the rest of the dataset. This process may be costly and time-consuming. On the other hand, using pre-trained classifiers saves a lot of time. These models are easy-to-use with a couple of lines of code, but the specificity of their training datasets might constrain them.

Josh and Marshal suggested that pre-trained models for NLP should be tried out too. There are models readily available to use in the market that can predict the sentiment of the given data. Before implementing the models that the company has built, Robin asked you to collaborate with Josh and Marshal to experiment with the pre-trained model.

Some of the pre-trained models available are Vader, Happy Transformer, TextBlob, and Google cloud NL API. Marshal requested you go through the documentation of these models. The implementation of these models is pretty straightforward and simple. Some of them are open source and others are not so you should consider the cost before adopting them.

This week, you need to

- Identify the open source and paid services for the pre-trained model.
- Make a report comparing the cost and time estimates for building your own model and using the pre-trained model.

### Deliverables for This Week's Case Study

- Use any two open-source pre-trained models to predict the sentiment on airline reviews.
- Compare the accuracy and precision of your model and the pre-trained model.
- Analyze the complexity and speed of your model and pre-trained model and log it

---

**Program Name:** Big Data Analytics

**Project Code:** CPL-5559-DSMM

---

## Week 13

### Applicable VLOs or EES for This Week's Case Study

- Identify and assess data analytics and Big Data business strategies and workflows to respond to new opportunities or provide project solutions.
- Deliver data-oriented projects using data science, business analysis, and project management principles, tools, and techniques.

### This Week's Detailed Case Study Information

#### Submission of Project Report + Practice Presentation

- Finish Project Report for submission your final submission is due this week. Be proud of the work you have completed in this project, now you can spend time polishing your presentation and making sure you will capture the stakeholder's attention in a positive way.
- Review APA Guidelines and ensure your project has followed them. This is particularly important.

#### Hone your presentation skills.

- A Presentation for your Sentiment analysis project is meant to highlight your research findings and the conclusions, opportunities, and best practices that you can be followed on other projects. The analysis of your findings is one of the most important parts and should be conveyed in your presentation.

### Deliverables for This Week's Case Study

1. Final Project Report – this is your final document with all supporting resources: including any appendices. Bibliography and reference in APA format required.
2. Feedback Video
  - Prepare to answer questions regarding the project on client expectations, Job Market, and on how you will sell your product



**Program Name:** Big Data Analytics

**Project Code:** CPL-5559-DSMM

---

## Week 14

### Preparing for Your Final Week Activities

It is the end of your work term. Your supervisor is grateful for your efforts. The final week contains activities which include both individual and teamwork efforts. Take this opportunity to shine bright in the final activities.

### Final Week Deliverables and Format Requirements

Your supervisor will provide you with more detail about the Final Week responsibilities.

### AIP Project Completion

Following completion of the Final Week activities, you will be notified by your supervisor if you pass or fail the WIL Project.

**Program Name:** Big Data Analytics

**Project Code:** CPL-5559-DSMM

---

## Appendix

### Acronym Used

HDFS: Hadoop Distributed File system

PCA: Principal Component Analysis

NLP: Natural Language processing

PD: Personal Development

HQL: Hive query Language

SQL: Structured Query Language

RBAC: Role-Based Access Control

ML: Machine Learning

DE: Data Engineer

PM: Project Manager

PII: Personally identifiable information

AI: Artificial Intelligence

ITTO: Input Tools& Technologies and output.