**Research Assignment: Introduction to Machine Learning?**

**Define Machine Learning using a real-life example.**

ML is the process of training a piece of software, called a model, to make useful predictions or generate content (like text, images, audio, or video) from data.

**Compare Supervised Learning and Unsupervised Learning, giving an example of each?**

**Supervised learning** models can make predictions after seeing lots of data with the correct answers and then discovering the connections between the elements in the data that produce the correct answers. This is like a student learning new material by studying old exams that contain both questions and answers. Once the student has trained on enough old exams, the student is well prepared to take a new exam. These ML systems are "supervised" in the sense that a human gives the ML system data with the known correct results.

**Unsupervised learning** models make predictions by being given data that does not contain any correct answers. An unsupervised learning model's goal is to identify meaningful patterns among the data. In other words, the model has no hints on how to categorize each piece of data, but instead it must infer its own rules.

For example Unsupervised learning can be used to categorize a social media feed by identifying hidden patterns in user behavior and content. Instead of using predefined labels, the algorithm automatically groups users with similar interests or friends, or clusters content with similar themes, hashtags, and engagement metrics. This allows the social media platform to create personalized feeds without being explicitly told what a user likes, making it easier to discover new content and connect with new people.

**What causes Overfitting? How can it be prevented?**

**Overfitting** is an undesirable machine learning behavior that occurs when the machine learning model gives accurate predictions for training data but not for new data.Consider a use case where a machine learning model has to analyze photos and identify the ones that contain dogs in them. If the machine learning model was trained on a data set that contained majority photos showing dogs outside in parks , it may may learn to use grass as a feature for classification, and may not recognize a dog inside a room.

**To prevent overfitting**, increase the quantity and diversity of training data through data augmentation or collecting more data, simplify model complexity by removing features or using smaller models, apply regularization techniques

**Explain how training data and test data are split, and why this process is necessary?**

**We split data into a training set and a test set** to check if a computer has truly learned or just memorized.its like preparing for exam: the training set (most of the data) is your textbook, which you study and learn from. The test set (a smaller, separate portion) is your final exam, which is kept secret until the day of the eaxam. This split is necessary because if you tested the computer on the exact same data it learned from, it would get a perfect score by memorization alone, which doesn't prove it can handle new, real-world problems. By testing it on unseen data, we get an honest grade on how well it actually learned the general patterns.

**Find one case study (research paper or article) that explains how Machine Learning has been applied in healthcare, business, or transportation. Summarize its findings?**

machine learning is being applied in healthcare to significantly improve patient care and data management. It helps medical professionals analyze vast amounts of patient data to identify health trends and predict treatment outcomes. Key applications include using neural networks to analyze medical images, natural language processing to extract data from doctor's notes, and robotics to assist in surgeries. The article provides real-world examples like disease prediction, personalized treatment plans, and new medication development. While the technology holds great promise, the article also notes important ethical considerations, such as patient privacy and the potential for biased algorithms, which must be carefully addressed.

[Reference](http://coursera.org/articles/machine-learning-in-health-care)