# How to pursue career in machine learning

# Prerequisites to learn ML

## **Statistics and probability:** You should have a strong understanding of the statistical concepts such as probability, hypothesis testing, and estimation I order to understand how machine learning algorithm works.

**Linear algebra:** Linear algebra is the study of vector spaces and linear transformations. It is a fundamental tool for machine learning, as many machine learning algorithms are based on linear algebra operations.

**Calculus:** Calculus is the study of change. It is used in machine learning to analyze the performance of algorithms and to optimize their parameters.

**Programming**: You need to be able to code in order to implement machine learning algorithms. The most popular programming languages for machine learning are Python and R.

**Data science:** Data science is the field of study that deals with the collection, cleaning, analysis, and interpretation of data. Machine learning is a subfield of data science, so you should have a basic understanding of data science concepts.

In addition to these technical prerequisites, you should also have the following soft skills:

**Problem-solving skills:** Machine learning is a problem-solving field. You should be able to identify problems, formulate solutions, and evaluate the results of your solutions.

**Critical thinking skills:** You should be able to think critically about data and algorithms. You should be able to identify potential problems and come up with creative solutions.

# Job Roles Related to ML in Industries

1. **Data Scientist:** Data scientists use their knowledge of statistics, machine learning, and programming to extract insights from data. They work with businesses to solve problems and make better decisions.
2. **Machine Learning Engineer:** Machine learning engineers build and deploy machine learning models to solve real-world problems .They work with data scientists to collect, clean, and prepare data, and with software engineers to deploy models into production.
3. **Artificial Intelligence Engineer:** Artificial Intelligence Engineers are responsible for creating and maintaining AI systems that can learn and make decisions like humans. They use their knowledge of computer science, mathematics, and statistics to design and develop algorithms that can solve real-world problems.

1. **Business Intelligence Analyst:** Collects and analyzes data to identify trends and patterns that can help businesses make better decisions.
2. **Research Scientist:** A Research Scientist in the IT industry is responsible for conducting research and development of new IT products and services. They work on a variety of projects, from developing new algorithms to testing new hardware. They also collaborate with other engineers and scientists to bring new ideas to market.
3. **Computer Vision Engineer:** Develop and implement computer vision algorithms and systems that enable machines to interpret and understand the world around them.

Work with a variety of stakeholders, including software engineers, product managers, and business leaders, to ensure that their solutions meet the needs of the organization.

1. **Natural Language Processing (NLP) Engineer :** NLP engineers use their knowledge of computer science, statistics, and linguistics to develop systems that can understand and process human language.

# Tools and technologies used in ML:

**Programming languages**: Python, R, and Java are the most popular programming languages for machine learning.

**Machine learning frameworks**: TensorFlow, PyTorch, and scikit-learn are popular machine learning frameworks that provide a set of tools and libraries for developing and deploying machine learning models.

**Data science tools**: Jupyter Notebook, Pandas, and NumPy are popular data science tools that are used for data cleaning, analysis, and visualization.

**Cloud computing platforms:** Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform offer machine learning services that make it easy to build and deploy machine learning models.

**Version control systems**: Git and GitHub are popular version control systems that are used to track changes to code and data.

**Machine learning debuggers:** Weights & Biases and MLFlow are popular machine learning debuggers that help you track the performance of your machine learning models.

To find pre-collected datasets that are unbiased and relevant to our project, we can use websites such as Kaggle and GitHub. Many machine learning projects on GitHub include the dataset that they used, so this can be a great way to find high-quality data.

[**Kaggle**:](https://www.kaggle.com) Kaggle is a website that hosts a large collection of machine learning datasets. The datasets are contributed by users from all over the world, and they cover a wide range of topics. Kaggle also hosts competitions where users can compete to build the best machine-learning model on a given dataset.

[**Github:**](https://www.github.com)GitHub is a website where developers can host their code. Many machine learning projects are hosted on GitHub, and many of these projects include the datasets that they used. This can be a great way to find high-quality data that has already been cleaned and prepared for machine learning.

Some other websites like Kaggle where you can find pre-collected datasets for machine learning:

* [DrivenData](https://www.drivendata.org/)
* [Topcoder](https://www.topcoder.com/)
* [CrowdANALYTIX](https://www.crowdanalytix.com/)
* [TunedIT](https://tunedit.org/)

## References:

[1] <https://business.wfu.edu/masters-in-business-analytics-online/articles/machine-learning-job-roles/>

[2] <https://intellipaat.com/blog/prerequisites-for-machine-learning/>

[3] <https://www.analyticsvidhya.com/blog/2023/05/machine-learning-engineer-salary/>