# Intelligence Transpired through the Prognosis using Transfer Learning

**Preface**

**TABLE OF CONTENTS**

**1. SOFTWARE DEVELOPMENT LIFE CYCLE**

**2. PLATFORM KNOWLEDGE**

**3. DOMAIN KNOWLEDGE**

**4. ABOUT PROJECT**

4.1 ABSTRACT

4.2 SCOPE OF THE PROJECT

4.3 EXISTING SYSTEM

  4.3.1 DISADVANTAGES

     4.4 PROPOSED SYSTEM

  4.4.1 ADVANTAGES

**5. BOTTOM LINE AND FUTURE ENHANCEMENT**

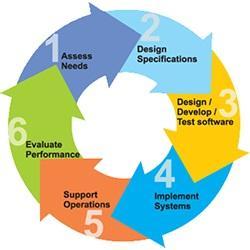
**6. HARDWARE AND SOFTWARE REQUIREMENTS**

**1. SDLC (Software Development Life Cycle)**

The Software Development Life Cycle is a systematic process for building software that ensures the quality and correctness of the software built. SDLC process aims to produce high-quality software which meets customer expectations. The software development should be completed within the pre-defined time frame and cost.

**SDLC Phases**

The entire SDLC process is divided into the following stages:



* Phase 1: Requirement collection and analysis
* Phase 2: Feasibility study
* Phase 3: Design
* Phase 4: Coding
* Phase 5: Testing
* Phase 6: Installation/Deployment
* Phase 7: Maintenance

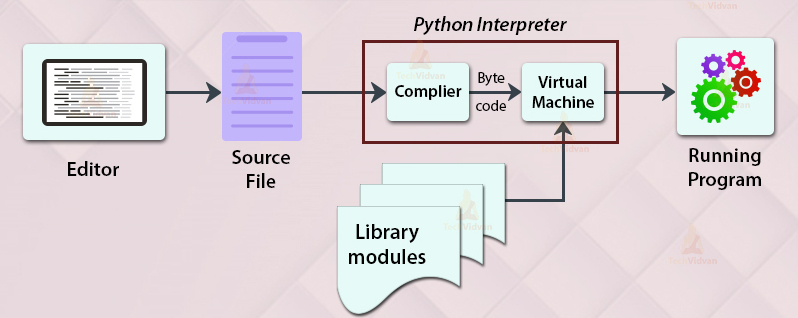
**2. Platform Knowledge**

**Introduction to Python:**

Python is developed by Guido van Rossum. Guido van Rossum started implementing Python in 1989. Python is a facile programming language so even if you are new to programming, you can learn python without facing any issues. Python is a general-purpose programming language that is becoming ever more popular for data science. Companies worldwide are using Python to harvest insights from their data and gain a competitive edge. Python specifically for data science. To store and manipulate data, and helpful data science tools to begin conducting your own analyses.

**What is Python?**

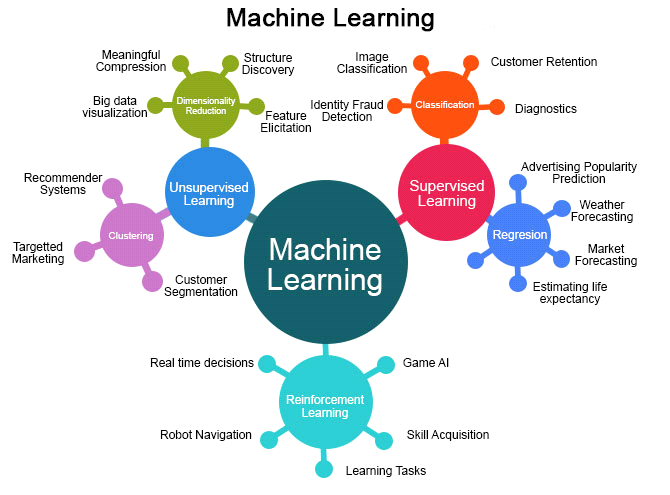
Python is an interpreted, high-level, general purpose programming language. It is dynamically typed and collected. Python is an interpreted language and not a compiled one, although compilation is a step. Python code, written in **.py** file is first compiled to what is called byte code which is stored with a **.**pyc or **.**pyo format. Instead of translating source code to machine code like C++, Python code it translated to byte code. This byte code is a low-level set of instructions that can be executed by an interpreter.  One popular advantage of interpreted languages is that they are platform-independent. As long as the Python byte code and the Virtual Machine have the same version, Python byte code can be executed on any platform (Windows, MacOS, etc).Dynamic typing is another advantage. In static-typed languages like C++, you have to declare the variable type and any discrepancy like adding a string and an integer is checked during compile time*.* In older programming languages, memory allocation was quite manual. Many times when you use variables that are no longer in use or referenced anywhere else in the program, they need to be cleaned from the memory. Garbage Collector does that for you.

****

**3. Domain Knowledge**

**OVERVIEW:**

Machine learning is an application of Artificial Intelligence (AI) that provides systems the ability to automatically learn and improve experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves. The primary aim is to allow the computer to learn automatically without human intervention or assistance and adjust actions accordingly. The test for a machine learning model is a validation error on new data not a theoretical test that proves a null hypothesis. Because this often uses an iterative approach to learn from data, the learning can be easily automated.



**MACHINE LEARNING METHODS**

Some of the methods of Machine Learning algorithm are categorized as

**SUPERVISED LEARNING**

A Supervised learning algorithm learns from labelled training data, helps you to predict outcomes for unforeseen data. It is highly accurate and trustworthy method.

**UNSUPERVISED LEARNING**

Unsupervised learning algorithm is the type of self - organized with the help of previously unknown patterns in dataset without pre-existing labels.

**SEMI-SUPERVISED LEARNING**

Semi-supervised learning is the combination of both supervised and unsupervised which means labelled and unlabelled data.

**REINFORCEMENT MACHINE LEARNING**

Reinforcement machine learning is an area of machine learning concerned with how software agents ought to take actions in an environment so as to maximize some notion of cumulative reward.

**APPLICATIONS OF MACHINE LEARNING**

* + Video Surveillance
  + Social Media Services
  + Email Spam and Malware Filtering
  + Financial Services
  + Health Care
  + Retail
  + Transportation

**ADVANTAGES**

* + Computational property is cheaper and more powerful.
  + Affordable data storage.
  + It can analysis complex data quickly and automatically.
  + It produces more accurate results.

**4About the project:**

**4.1 Abstract:**

In any organization Transfer Learning is a major boost for the company’s growth. This transfer learning helps in the research problem, mainly in the field of Machine Learning, which have been focused on strong knowledge gained while solving one problem and applying it to a different but related problem. Image captioning and question-answering are only two examples of the many issues to which transfer learning has been successfully used. Unfortunately, we may see substantial variance in performance in dynamic situations because the datasets we have used for training are frequently too large to completely explore all the variables of a problem configuration and their impact on the accuracy of the model. We can say an example for a car, knowledge gained while recognizing car; so that it can be different with truck. In the main model of any industry that too in industry like ONGC there may be the problem in the pipe which may result in full damage. From this there will be a log which helps in obtaining a trained dataset, from which we can obtain a particular day which helps in the organisation to plan according, in this case the transfer learning comes in handy for any organization.

**4.2 Scope of the project:**

Usually in any automation Industry usage of transfer learning and making it happens is too hectic, but in our proven analysing way it is more feasible to make it happen. The challenge of learning from data that was produced by a different generating process than that used for training has baffled many intelligent people for years. The most typical illustration of transfer learning is when the target learning task's output distribution or feature representation considerably differs from that of the source learning task. Because of the so said reasons many organization are being hesitant to switch to transfer learning and use its full utility to make it more appropriate for their work process. As we have said earlier in our ONGC Industrial application the process which takes more time is the maintenance work in case of any damages in the pipes, where the flushing should happen immediately and the days to happen should be predicted. Here in our model this happens very effectively and transfer learning takes its major step in prediction of date. By which the complexity of the transfer learning is reduced through our model.

**4.3 Existing system:**

Concept Drift is a frequent occurrence in transfer learning, particularly when the distributions used for training and testing are different. The domain adaptation technique and the idea drift-tolerant approach are the two basic mitigation strategies for this issue. In the existing paper, a novel concept drift-tolerant transfer learning algorithm that have be utilised to mitigate concept drift's detrimental effects on transfer learning. In concept drift-tolerant transfer learning, the outcome for unlabeled information is evaluated using data from a separate source. Most of the time, adding new features won't boost accuracy. Use similar labels between the training and test datasets while developing a prediction model with Concept Drift-Tolerant Transfer Learning.

**4.3.1 Disadvantages:**

* Selection of appropriate kernel function is a tedious operation and practicality in those functionalities is dimmed.
* Using a large data set requires more amount of time, which reduces the efficiency in the processing of data.
* We cannot predict the features and the outcome of the data, which leads to a wastage of time and cost-effectiveness.
* Even if the data changes are slight, the outcome will be a big difference.
* Those tools are complex and require training to use.

**4.4 Proposed system:**

Data in any sector is a tedious job that too in an automation industry where there is numerous amount of data to be processed, the delay which is about to happen is a difficult task to evade. But it is not so impossible to avoid, here in our model we have proposed the system which makes sure that our data which is to be processed are dimensionally good. By making sure that the mining of data, most of the insufficiencies which are about to happen are rectified. Then we use Passive Aggressive Regression algorithm which handles the prediction process in our application, this is the crucial part in which transfer learning happens in every process, which consumes a lot of energy. Because of that, the computational efficiency could go down if we don’t use the PAR algorithm. In our model, it is effective in terms of computational complexities and in lower time consumption.

**4.4.1 Advantages:**

* The passive aggressive regression approach is advantageous for our system, because it uses few resources.
* Large datasets can be handled via passive aggressive regression, which also modifies its model based on each new occurrence it meets.
* Using a passive aggressive regression model, more precise predictions can be made.
* This is most popular for big data applications.
* Large-scale learning typically uses passive-aggressive algorithms.

**5. BOTTOM LINE AND FUTURE ENHANCEMENT:**

Therefore, in past usage of the prediction, there is difficulty in getting data. Thus in our proposed model, we achieve it by transfer learning, so that we get better results in terms of computational productivity with better time efficiency. Next we predict our exact maintenance time through our passive aggressive regression which predicts the value. However, as mentioned Data mining of this kind enables computers to "learn" on their own through data analysis and pattern detection, but the effectiveness is not fulfilled. These exceptions will be solved in the future.

**6. HARDWARE AND SOFTWARE REQUIREMENTS**

**Hardware requirements:**

* Processor: Intel (R) Pentium (R)
* Speed: 1.6 GHz and Above
* RAM: 6 GB and Above
* Hard Disk: 120 GB
* Monitor: 15’’ LED SVGA
* Input Devices: Keyboard, Mouse

**Software requirements:**

* Operating system: Windows 10
* Coding Language: PYTHON
* IDE: PyCharm
* Database: MySQL