# PART 1:

# INTRODUCTION TO IDS INFOTECH:

****

Established in 1989, IDS Infotech Ltd. (IDS), has been a preferred Business Process Management and Software Solution provider to Fortune 500 companies and AM100 law firms worldwide for over two decades. A team of over 1,000 trained and skilled professionals, work round-the-clock to deliver quality solutions such as patent research and analytics, technology solutions, content management, engineering design and manufacturing support, physician practice management, eDiscovery and paralegal support, technical help-desk and other industry specific back-office processes. With various delivery centers in North India, IDS has multiple front-end offices globally, including US, UK and the Netherlands.

*Legal Services-* [IDS- Legal has been a worldwide leader in providing legal support services since 1999. We provide Fortune 1000 companies and AM250 Law Firms with an elite group of professionals providing e-Discovery and Paralegal services.](http://www.ids-legal.com/)

*Engineering Services-* [Inde Dutch Engineering & Aerospace Services Limited ("IDEAS"), an AS9100 Rev D certified company, is a leading engineering solutions provider catering to customers in the Aerospace, Semiconductors, Automotive, Machinery, and Affiliated Industry segments.](http://www.ideasengg.com/)

*Software Services-* [With a team of 100+ Software Engineers, IDS provides its customers with customized software solutions, supporting the end to end cycle of product development. We offer an exceedingly productive fusion of business and technological expertise utilizing our in-house domain experts.](http://ids-technologies.in/)

*Content Publishing And Management Service -* [The Content Management and Publishing Services Group (CMPSG) at IDS Infotech Ltd., specializes in delivering customized, technology driven data capture, data conversion, e-publishing, pre press and editorial solutions.](http://cmp.idsil.com/)

*Patent Research And Analytics-* [IDS-IP offers a suite of Patent Research and Analytics Services in creation, protection and monetization of patent rights. Our processes are specially designed and tailored to provide patent protection.](http://www.ids-ip.com/)

*HealthCare Services-* [IDS-Healthcare is a specialized team of professionals providing a gamut of Healthcare services to Hospitals, CROs, Medical Management Companies, Insurance providers, etc. worldwide.](http://www.idsargus.com/)

*Environment Health And Safety Services-* [IDS has been providing (M)SDS services to Environment, Health and Safety (EH&S) compliance and information management services companies in USA and Canada for over eight years.](http://idsil.com/environment.php)

*Medical Scribe-* [At IDS, we offer virtual scribe services to US doctors with marvels of technology to stream live patient visits to our HIPAA compliant facilities , whereby IDS Medical Scribes undertake real time updates of US hospital EHRs.](https://medicalscribe.idsil.com/)

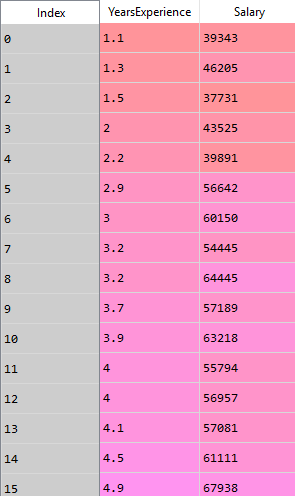
# PART 2:

# SUPERVISED MACHINE LEARNING:

**1). Regression:**

a). Simple Linear Regression- Simple linear regression is useful for finding relationship between two continuous variables. One is predictor or independent variable and other is response or dependent variable.

* Example:



(Years of Experience vs Salary dataset)



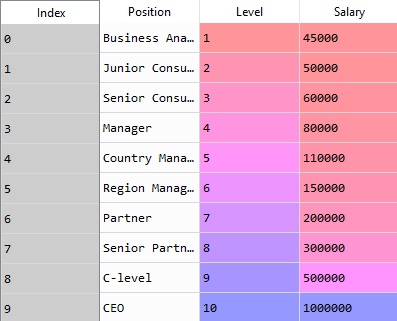
(Plotting the training set)

b). Polynomial Regression- Polynomial regression is a form of regression analysis in which the relationship between the independent variable x and the dependent variable y is modelled as a nth degree polynomial in x.

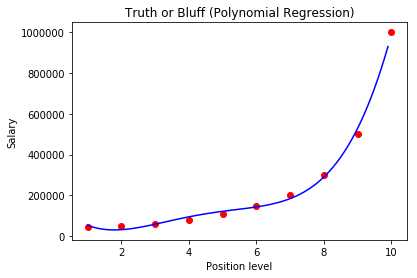
The kth order polynomial model in one variable is given by

y=a0 + a1\*x + a2\*x\*x + a3\*x\*x\*x + ……… (ak)\*(x^k)

Consider the dataset given below.



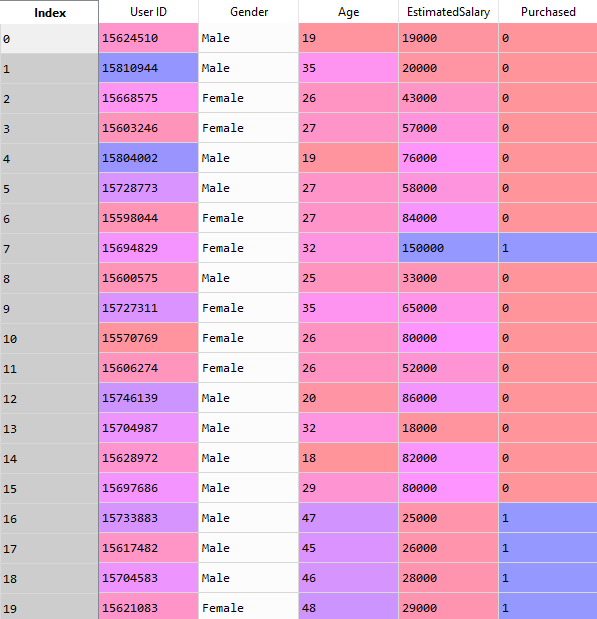
Curve fits the dataset very precisely here in the case of polynomial regression.



**2). Classification:**

a). Logistic Regression- It is the go-to method for binary classification problems (problems with two class values).

The two possible dependent variable values are often labelled as "0" and "1", which represent outcomes such as pass/fail, win/lose, alive/dead or healthy/sick.

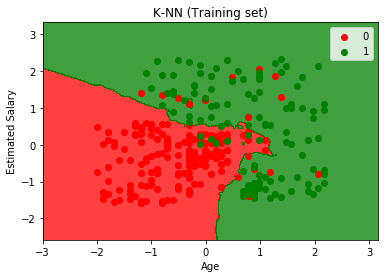
The above dataset is used to make the machine learn and thus to predict whether a customer is going to buy the product or not based on the variables such as his/her salary, age etc.



So, two separate regions got classified. However not all the predictions are correct as some of the green points are in the red region while some of the red points are in green region. Still, it’s a good model to predict whether the customer is going to buy the product or not.

b). K-Nearest Neighbors(K-NN)-  **KNN** is a **non-parametric, lazy**learning algorithm. Its purpose is to use a database in which the data points are separated into several classes to predict the classification of a new sample point.

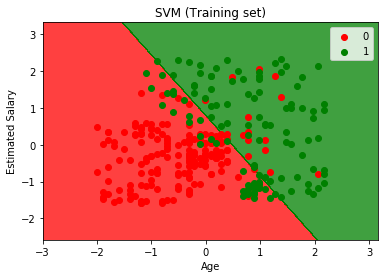
We are going to plot the previous dataset using this algorithm this time.

****

This time the boundary is not a straight line as it was in logistic regression and therefore it is more accurate in predicting the dependent variable.

c). Support Vector Machine- A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples. In two-dimensional space this hyperplane is a line dividing a plane in two parts where in each class lay in either side.

We are again using the earlier dataset to see how this algorithm predicts or performs in comparison to others.



It also makes a line dividing the plane in two regions just like the logistic regression. However, in the case of logistic regression the decision boundary needs not to be a straight line always. Its shape can vary.

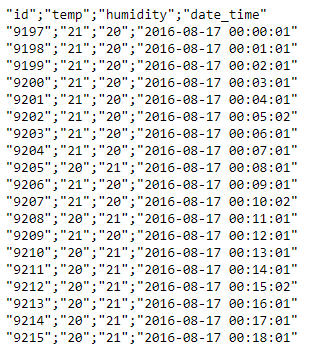
# PART 3:

# ANALYSIS OF TEMPERATURE AND HUMIDITY AT COMPANY’S OFFICE OF LAST TWO YEARS:

**INTRODUCTION**:

We have been shared a file that contains each day’s temperature and humidity entries at the company’s office for the last 2 years and we are supposed to find or establish a relationship between the temperature/humidity and the time of the day it was recorded at. Here is a sample of the dataset we had been given. It has the following variable as columns:

1. ID
2. Temperature
3. Humidity
4. Date and time



**PROGRESS**:

Till now, we have done the data preprocessing part completely. i.e. we made a table with the above listed variables as its columns.

The next thing we need to do is to make a model using one of the various machine learning algorithms so as to find the relationship if there exists any. Attached below is the preprocessed data in the form of a table.

