# Topic: Support Vector Machines (SVM)

**Instructions:**

Please share your answers filled in-line in the word document. Submit code separately wherever applicable.

Please ensure you update all the details:

**Name:RAKESH SETHU NP Batch ID: 11052022\_7.30PM**

**Topic: SVM**

**Hints:**

1. **Business Problem**
   1. **What is the business objective?**
   2. **Are there any constraints?**
2. **Work on each feature of the dataset to create a data dictionary as displayed in the below image:**



**2.1 Make a table as shown above and provide information about the features such as its data type and its relevance to the model building. And if not relevant, provide reasons and a description of the feature.**

**3.Data Pre-processing**

**3.1 Data Cleaning, Feature Engineering, etc.**

**3.2 Outlier Treatment.**

**4. Exploratory Data Analysis (EDA):**

**4.1 Summary.**

**4.2 Univariate analysis.**

**4.3 Bivariate analysis.**

**5. Model Building**

* 1. **Build the model on the scaled data (try multiple options)**
  2. **Use the SVM algorithm.**
  3. **Train and test the model and compare accuracies by building a confusion matrix and use different hyperparameters.**
  4. **Briefly explain the model output in the documentation.**

**6. Write about the benefits/impact of the solution - in what way does the business (client) benefit from the solution provided?**

**Problem Statement: -**

A construction firm wants to develop a suburban locality with new infrastructure but they might incur losses if they cannot sell the properties. To overcome this, they consult an analytics firm to get insights on how densely the area is populated and the income levels of residents. Use the Support Vector Machines algorithm on the given dataset and draw out insights and also comment on the viability of investing in that area.

**Objective:** Maximize the accuracy in predicting using svm which identify optimal separating hyperplane.

**Constraint:** Minimize error in choosing a good kernel.

|  |  |  |  |
| --- | --- | --- | --- |
| **Name of the Feature** | **Type** | **Description** | **Relevance** |
| age | Discrete | Age of a person | Relevant, Provides useful information. |
| workclass | Qualitative, Ordinal | Work class of person | Relevant, Provides useful information. |
| education | Qualitative, Ordinal | Education of a person | Relevant, Provides useful information. |
| educationno | Nominal | Number given to education | Irrelevant, does not Provides useful information. |
| maritalstatus | Qualitative, Ordinal | Marital status of a person | Relevant, Provides useful information. |
| occupation | Qualitative, Ordinal | occupation of a person | Relevant, Provides useful information. |
| relationship | Qualitative, Ordinal | relationship of a person | Relevant, Provides useful information. |
| race | Qualitative, Ordinal | race of a person | Relevant, Provides useful information. |
| sex | Qualitative, Ordinal | Gender of a person | Relevant, Provides useful information. |
| capitalgain | Quantitative, Ratio | Profit of a person | Relevant, Provides useful information. |
| capitalloss | Quantitative, Ratio | Loss of a person | Relevant, Provides useful information. |
| hoursperweek | Quantitative, discrete | Hours per week worked | Relevant, Provides useful information. |
| native | Nominal | Native of a person | Irrelevant, does not Provides useful information. |
| Salary | Quantitative, discrete | Salary of the person | Relevant, Provides useful information. |



**Problem Statement: -**

In California, annual forest fires can cause huge loss of wildlife, human life, and can cost billions of dollars in property damage. Local officials would like to predict the size of the burnt area in forest fires annually so that they can be better prepared in future calamities.

Build a Support Vector Machines algorithm on the dataset and share your insights on it in the documentation.

Note: - Size\_ category is the output variable.

**Objective:** Maximize the accuracy in predicting using svm which identify optimal separating hyperplane.

**Constraint:** Minimize error in choosing a good kernel.

|  |  |  |  |
| --- | --- | --- | --- |
| **Name of the Feature** | **Type** | **Description** | **Relevance** |
| month | Ordinal | Month in which fire occurs | Relevant, Provides useful information. |
| day | Ordinal | Day of the week fire occurs | Relevant, Provides useful information. |
| FFMC | ratio | Fine fuel moisture code | Relevant, Provides useful information. |
| DMC | ratio | Duff moisture code | Relevant, Provides useful information. |
| Dc | ratio | Drought code | Relevant, Provides useful information. |
| ISI | ratio | Initial spread index | Relevant, Provides useful information. |
| Temp | ratio | Temperature of fire | Relevant, Provides useful information. |
| RH | ratio | Relative humidity | Relevant, Provides useful information. |
| Wind | ratio | Wind speed | Relevant, Provides useful information. |
| Rain | ratio | Rain scale | Relevant, Provides useful information. |
| Area | ratio | Area in which fire burns | Relevant, Provides useful information. |

