**Use Case 1: Create a feature vector to categorise various songs into different genres**

**Introduction:** The client is an online music streaming platform that publishes music from various artists across the globe. This company is looking to develop a machine learning algorithm that can automatically classify various music uploaded by various artists who are part of the companies accredited publishers into pre-defined genres.

**Aim of the Assignment:** Aim of this assignment is to come up with a feature vector that list all the various attributes of the songs; which will be considered for classification of the song into a specific genre.

**List of Feature’s Considered:** Below list defines various features that can be considered for classification of a song in a specific genre. This is based on the metadata of the song and the actual composition of the song:

**Primary Classification: -**

|  |  |  |
| --- | --- | --- |
| **Song Level Feature Vector (VSongMD)** | | |
| Feature 1 | Tempo | Based on a pre-defined scale of tempo songs can be classified into specific genre |
| Feature 2 | Treble | Based on a pre-defined scale of treble songs can be classified into specific genre |
| Feature 3 | Bass | Based on a pre-defined scale of bass songs can be classified into specific genre |
| Feature 4 | Overall Average Frequency | Based on a pre-defined scale of average frequency songs can be classified into specific genre |
| Feature 5 | Peak Frequency | Based on a pre-defined scale of pear frequency songs can be classified into specific genre |
| Feature 6 | Trough Frequency | Based on a pre-defined scale of trough frequency songs can be classified into specific genre |
| Feature 7 | Song Decibel Level | Based on a pre-defined scale of decibel level songs can be classified into specific genre |

We can use the **VSongMD** as a primary classification feature vector. This can be normalized by additional set of features listed below

**Normalization Classification: -**

|  |  |  |  |
| --- | --- | --- | --- |
| **Metadata Level Feature Vector (VSong)** | | | |
| Feature 8 | Singer | Based on a pre classified list of singer genre a song from a specified singe can be classified |
| Feature 9 | Band Name | Based on a pre classified list of band genre a song from a specified singe can be classified |
| Feature 10 | Year of Release | Based on the year of release songs can be classified into a set of genres |
| Feature 11 | Song Duration | Duration of a song could provide some indication on the genre of the song based on a pre-defined scale |

**Use Case 2: Create a feature vector to categorise Xray images into Covid-19 affected and normal Bacterial Pneumonia**

**Introduction:** Amides the COVID-19 pandemic the number of cases related to lung infections have drastically increased and the healthcare workers around the globe are overwhelmed by the shire number of cases that needs to be attended on a priority basis. Hence there is a urgent need to develop a algorithm that can systemically classify vast number of Xray images of the lungs into COVID-19 related and Bacterial Pneumonia related catagories so that patients who are affected with COVID-19 can be treated in special facilities with minimal lead time.

**Aim of the Assignment:** Aim of this assignment is to come up with a feature vector that list all the various attributes of the lung Xray; which will be considered for classification of the xRay into COVID-19 or Bacterial Pneumonia related lung dieses.

**List of Feature’s Considered:** Below list defines various features that can be considered for classification of lung xRay. This is based on patient history, sex, pre disposed conditions and actual anatomy of the organ:

**Primary Classification: -**

|  |  |  |
| --- | --- | --- |
| **Anatomy Level Feature Vector (VAnatomy)** | | |
| Feature 1 | Length (Lobe 1) Right | The Xray dimensions can be compared to the clinical dimensions of the healthy lung to identify amount of variance |
| Feature 2 | Breath (Lobe 1) Right | The Xray dimensions can be compared to the clinical dimensions of the healthy lung to identify amount of variance |
| Feature 3 | Length (Lobe 2) Right | The Xray dimensions can be compared to the clinical dimensions of the healthy lung to identify amount of variance |
| Feature 4 | Breath (Lobe 2) Right | The Xray dimensions can be compared to the clinical dimensions of the healthy lung to identify amount of variance |
| Feature 5 | Length (Lobe 3) Right | The Xray dimensions can be compared to the clinical dimensions of the healthy lung to identify amount of variance |
| Feature 6 | Breath (Lobe 3) Right | The Xray dimensions can be compared to the clinical dimensions of the healthy lung to identify amount of variance |
| Feature 7 | Length (Lobe 1) Left | The Xray dimensions can be compared to the clinical dimensions of the healthy lung to identify amount of variance |
| Feature 8 | Breath (Lobe 1) Left | The Xray dimensions can be compared to the clinical dimensions of the healthy lung to identify amount of variance |
| Feature 9 | Length (Lobe 2) Left | The Xray dimensions can be compared to the clinical dimensions of the healthy lung to identify amount of variance |
| Feature 10 | Breath (Lobe 2) Left | The Xray dimensions can be compared to the clinical dimensions of the healthy lung to identify amount of variance |
| Feature 11 | Length (Overall) Right | The Xray dimensions can be compared to the clinical dimensions of the healthy lung to identify amount of variance |
| Feature 12 | Breath (Overall) Left | The Xray dimensions can be compared to the clinical dimensions of the healthy lung to identify amount of variance |
| Feature 13 | Diameter of Limp nodes | This can be used to differentiate between bacteria, virus or cancer related developments |
| Feature 14 | Diameter of Dark Exposure on X Ray (Features) | This can provide view of amount of water retention that can be compared to clinical standards for virus or bacterial related infections |
| Feature 15 | Diameter of Light Exposure on X Ray (Features) | This can provide view of amount of water retention that can be compared to clinical standards for virus or bacterial related infections |

We can use the **VAnatomy** as a primary classification feature vector. This can be normalized by additional set of features listed below

**Normalization Classification: -**

|  |  |  |  |
| --- | --- | --- | --- |
| **Patent Information Level Feature Vector (VPINFO)** | | | |
| Feature 16 | Sex | A secondary One Hot Encoding vector can be used to identify between Male, Female or undeclared |
| Feature 9 | Age | A secondary One Hot Encoding vector can be used to group patients into age bands and allocate positions on a 0 -1 scale |
| Feature 10 | Previous History | A secondary One Hot Encoding vector can be used to group patients into previous history bands and allocate positions on a 0 -1 scale |
| Feature 11 | Travel Frequency | A secondary One Hot Encoding vector can be used to group patients into high, medium, low travel frequency bands and allocate positions on a 0 -1 scale |

The above feature vector can be used as a input to develop the required classification algorithms for the two identified use cases.