**Model Development Phase**

|  |  |
| --- | --- |
| Date | 12 July 2024 |
| Team ID | SWTID1720077079 |
| Project Title | Wild Blueberry Yield Prediction |
| Maximum Marks | 5 Marks |

**Feature Selection Report**

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature** | **Description** | **Selected (Yes/No)** | **Reasoning** |
| Clonesize | The average blueberry clone size in the field | Yes | Univariate, bivariate, multivariate analysis shows good correlation |
| Honeybee | Honeybee density in the field | Yes | Univariate, bivariate, multivariate analysis shows good correlation |
| Bumbles | Bumblebee density in the field | Yes | Univariate, bivariate, multivariate analysis shows good correlation |
| Andrena | Andrena bee density in the field | Yes | Univariate, bivariate, multivariate analysis shows good correlation |
| Osmia | Osmia bee density in the field | Yes | Univariate, bivariate, multivariate analysis shows good correlation |
| MaxOfUpperTRange | The highest record of the upper band daily air temperature during the bloom season | Yes | Univariate, bivariate, multivariate analysis shows good correlation |
| MinOfUpperTRange | The lowest record of the upper band daily air temperature | Yes | Univariate, bivariate, multivariate analysis shows good correlation |
| AverageOfUpperTRange | The average of the upper band daily air temperature | Yes | Univariate, bivariate, multivariate analysis shows good correlation |
| MaxOfLowerTRange | The highest record of the lower band daily air temperature | Yes | Univariate, bivariate, multivariate analysis shows good correlation |
| MinOfLowerTRange | The lowest record of the lower band daily air temperature | Yes | Univariate, bivariate, multivariate analysis shows good correlation |
| AverageOfLowerTRange | The average of the lower band daily air temperature | Yes | Univariate, bivariate, multivariate analysis shows good correlation |
| RainingDays | The total number of days during the bloom season, each of which has precipitation larger than zero | Yes | Univariate, bivariate, multivariate analysis shows good correlation |
| AverageRainingDays | The average of raining days of the entire bloom season | Yes | Univariate, bivariate, multivariate analysis shows good correlation |
| fruitset | Refers to the proportion of flowers that develop into fruits. | No | Shows very low correlation with other features from multivariate analysis |
| fruitmass | Indicates the mass (weight) of the fruits. | No | Shows very low correlation with other features from multivariate analysis |
| seeds | Represents the number of seeds produced within the fruits. | No | Shows very low correlation with other features from multivariate analysis |