**Model Development Phase Template**

|  |  |
| --- | --- |
| Date | 25 March 2025 |
| Team ID | SWTID1749753590 |
| Project Title | Early prediction for chronic kidney disease detection: a progressive approach to health |
| Maximum Marks | 5 Marks |

**Feature Selection Report Template**

In the forthcoming update, each feature will be accompanied by a brief description. Users will indicate whether it's selected or not, providing reasoning for their decision. This process will streamline decision-making and enhance transparency in feature selection.

| **Feature** | **Description** | **Selected (Yes/No)** | **Reasoning** |
| --- | --- | --- | --- |
| **Hemoglobin** | **Protein in red blood cells; low levels indicate anemia and CKD** | **Yes** | **Strongly correlated with CKD. Selected for high predictive value.** |
| **Packed Cell Volume** | **Volume % of red blood cells in blood** | **Yes** | **Indicates anemia, a common CKD symptom.** |
| **Specific Gravity** | **Urine concentration level** | **Yes** | **Reflects kidney's ability to concentrate urine. Important diagnostic metric.** |
| **Red Blood Cell Count** | **Number of RBCs in blood** | **Yes** | **Lower counts can signal kidney failure-related anemia.** |
| **Hypertension** | **Whether the person has high blood pressure** | **Yes** | **Major CKD risk factor; strongly associated.** |
| **Diabetes Mellitus** | **Presence of diabetes** | **Yes** | **Strong CKD cause; included due to its high clinical relevance.** |
| **Albumin** | **Protein content in urine** | **Yes** | **High levels suggest kidney damage. Critical feature for diagnosis.** |
| **Blood Glucose Random** | **Random blood sugar level** | **Yes** | **Linked to diabetes, which causes CKD.** |
| **Age** | **Age of the patient** | **No** | **Not selected due to lower statistical impact in model accuracy.** |
| **Blood Urea** | **Amount of urea in blood** | **No** | **Removed during preprocessing due to high correlation/redundancy.** |
| **Serum Creatinine** | **Waste in blood filtered by kidneys** | **No** | **Excluded to avoid data leakage since it's often a direct diagnosis marker.** |
| **Sodium** | **Sodium level in blood** | **No** | **Some missing values; lesser contribution during feature importance check.** |
| **Potassium** | **Potassium level in blood** | **No** | **Similar to sodium; excluded due to missing data and low model impact.** |
| **White Blood Cell Count** | **WBC count in blood** | **No** | **Had missing values and less predictive power.** |
| **Appetite** | **Normal or poor appetite** | **No** | **Categorical feature; excluded during initial model simplification.** |
| **Pedal Edema** | **Swelling in feet** | **No** | **Categorical and sparsely distributed; excluded for simplicity.** |
| **Anemia** | **Whether the patient is anemic** | **No** | **Highly correlated with hemoglobin; redundant information.** |
| **Serum Albumin** | **Albumin in the blood** | **No** | **Overlapping with urine albumin; not added to avoid redundancy.** |
| **Pus Cell** | **Number of pus cells in urine** | **No** | **Categorical and incomplete in dataset.** |
| **Bacteria** | **Presence of bacteria in urine** | **No** | **Sparse binary feature; dropped due to limited contribution.** |