**INCOMPLETE CASE K NEAREST NEIGHBOUR IMPUTATION [ICKNNI] AND COMPLETE CASE K NEAREST NEIGHBOUR IMPUTATION [CCKNNI]**

Arulpiruthiviraj Arunthavaraja (105212573), Baranidharan Pasupathi (110007279) and Sariat Sultana (110028032)

**ABSTRACT –** The Project aims to evaluate the imputed missingness of ICKNNI and CCKNNI. We identified that including some incomplete instances for imputation would provide persuasive unbiased solutions for analysis. This initial report briefly explains the steps and complexity computation of both imputation techniques.

**NOTATION USAGE -** Following standard notations i.e., (input dataset ; instances ; variables &similarly &; missing variables ; observed variables . Let set of complete instances , where is complete instances; set of incomplete instances , where is incomplete instances; ; shortest distance between modified/output dataset .

1. **MAIN STEPS OF COMPLETE CASE K NEAREST NEIGHBOUR IMPUTATION**

**Step-1:** Initialize from and load it to i.e., [1].

**Step-2:** Initialize to and find

; and are observed variables of incomplete and complete instances; ; represents number of observed variables in a instance;

**Step-3:** Find complete nearest neighbour ;

**Step-4:** For ; is number of missing variables in an instance; imputes missing values;

**Step-5:** Imputed values and observed are stored in new instance

;

**Step-6:** is then equated to the variable

since . Finally,

1. **MAIN STEPS OF INCOMPLETE CASE K NEAREST NEIGHBOUR IMPUTATION**

**Step-1:** Initialize from and load it to i.e., ; Compute ; [2]

**Stpe-2:** Compute ; is a set of observed values of all instance. ; is a library that

holds all possible neighbour instances of , later their distances is computed.

**Step-3:** Since holds all instances obviously ; Find

**Step-4:** Find complete nearest neighbour ;

**Step-5:** predicts values of missing one;[1]

**Step-6:** Step-5 and Step-6 of Complete case K nearest neighbour is followed to derive complete dataset.

1. **COMPLEXITY ANALYSIS**

Considering dataset X with records & attributes. The computational complexity of KNN impute method is approximately , assuming . The execution of both cases will be almost similar since they follow KNN procedure. In ICKNNI where maximum percent of missing cases are considered there might be a slightly longer execution is expected while in CCKNNI it’s not. In-reality software empirical analysts use list deletion LD in CCKNNI datasets to avoid biased output which will make the execution time much faster when compared with ICKNNI.

**REFERENCE**

1. J. V. Hulse, T. M. Khoshgoftaar, “Incomplete-Case Nearest Neighbor Imputation In Software Measurement Data,” An International Journal, Computer and Electrical Engineering, Florida Atlantic University, Florida, USA, 2011.
2. Z. G. Liu, Q. Pan, J. Dezert, A. Martin, “Adaptive Imputation Of Missing Values For Incomplete Pattern Classification,” Pattern Recognition, Elsevier, 2016, 52, ff10.1016/j.patcog.2015.10.001ff. ffhal-01259203f.