

Indiscriminant Expected Maximization Imputation Model using Multiple Classification Technique on diabetic Dataset

M.Sangeetha, M. Senthil Kumaran



Abstract: *The missing data is inescapable in clinical research. While individuals with their left out or missing data may balance out when regards to those with no missing data to the extent the aftereffect interest in gauge all around. Missing data is of three types :Missing at random (MAR),Missing completely at random (MCAR), and Missing not at random (MNAR). In a medicalstudy, missing data is, to a great extent, MCAR. Missing information can build up deep troubles in the assessments and perception of results and undermine the authenticity of results to finish. Various strategies have been created for managing missing information. These incorporate total case examinations, missing pointer technique, single worth imputation, and affectability investigations were joining most pessimistic scenario while greate-case situations. Whenever connected the MCAR suspicion, a portion of these techniques can give unprejudiced, however frequently less exact assessments. Single value imputation an elective technique to manage missing information, which records for the vulnerability related to missing information. Single value imputation is actualized in most factual programming under the MAR suspicion and gives fair and substantial evaluations of affiliations dependent on data from the accessible information. The technique influences not just the coefficient gauges for factors with missing information, yet additionally the appraisals for different factors with zero missing information.*

Keywords: *Expected Maximization, performance analysis*

I. INTRODUCTION

Genuine information collected often contain Missing Values (MVs). There are numerous explanations behind the presence of MVs, notably when manual information is used in the system, and the therapeutic report enables few ascribes to be left out of the system when the clinical information is gathered as a major aspect of a clinical preliminary. This might be, on the grounds that the patient may not want certain data to be recorded about them (for example records of cardiac or melancholy) or on the grounds that they are not viewed as proper for the class of disease being dealt with. Exhibiting information examination becomes troublesome due to the MVs.

Revised Manuscript Received on August 30, 2019.

* Correspondence Author

M.Sangeetha*, SCSVMV University/Research Scholar, Enathur, Kanchipuram, 631561, India. SRM Institute of Science and Technology/Information Technology, Kantankulathur, Kancheepuram, India.

M. Senthil Kumaran, Associate Professor, SCSVMV University/Department of Computer Science, Enathur, Kanchipuram, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

The nearness of MVs typically requires a pre-handling venture before proceeding onward to the information extraction process. Unseemly treatment of the MVs can bring about deceiving or improper ends being drawn. The accompanying three kinds of issues are typically connected with MVs during information mining: loss of effectiveness; confusion in information taking care of and examination; and predisposition coming about because of contrasts among absent and complete information. [1] In the specific instance of order, fragmented information in either the preparation set or test set or two sets influence the expectation exactness. The reality of this issue depends to some degree on the extent of MVs.

The exploration announced in this paper expected to break down the utilization of a lot of attribution techniques on different classifiers. As a database, Diabetic Patients' dataset was utilized. The results will assist us in explaining how imputation might be a valuable apparatus to defeat the negative effect of MVs and to pick the most appropriate imputation arrangement blend for missing information through a program developed.

II. LITERARY SURVEY

Missing data regularly ordered into the accompanying three sorts: missing at random (MAR), missing completely at random (MCAR), and missing not at random (MNAR). [2] Right, when individuals with missing data are an unpredictable part of the examination people, the possibility of missing is alike in all cases; missing data are expressed as MCAR. An instance of MCAR is the time when a glass slide with biopsy information of a patient is coincidentally broken down to such a degree, that Pathol and histology examinations can't be executed, or when individuals had no heartbeat assessed as the equipment was broken. Therefore, MCAR with the missing data doesn't depend upon either observed data or indistinct data.

Rather than MCAR, the term MAR is unreasonable. Blemish happens when the missing or left data relies upon data we have observed. [3] For instance, the information in a low case study it is said to be MAR, while sexual orientation regarding men is more uncertain than ladies to round out the study. When sexual orientation is represented, the left out information does not rely upon the degree of their downturn. Another case of MAR is, an investigation of weight, information on good are more averse to be observed for more youthful people since they don't go to human services offices as frequently as more seasoned people.



At the point when the likelihood that information are missing relies upon the surreptitious information, for example, the estimation of the perception itself, at that point the missing information are signified as MNAR. [4] For instance, overweight or the less BMI people might be bound to their own weight estimated than people with ordinary weight, even though the age is represented. In this way, the explanation behind missingness is identified within secret qualities of the patient, and along these lines, information is MNAR. Another model is when people with extreme misery, or antagonistic impacts from stimulant medicine, are pretty much liable to finish a review on sadness. A third case is when information on pay is missing, while the likelihood of missing-ness is identified with the degree of salary, e.g., those with fluctuating pay won't report their pay.

Chart hypotheses have been useful in various disciplines in areas like arithmetic, building, software engineering, and science to decide or assess the systems of any missing attributes. [5] In epidemiological research, causal graphical models, for example, directed acyclic graphs (DAGs), can be utilized to decide if the information is MAR, MNAR or MCAR, along these lines illustrating the most fitting descriptive technique to manage missing qualities. [6]

III. PROPOSED WORK AND DESIGN

The Expected Maximization recipe is utilized to look out (local) most shot parameters of a connected math model in cases any place the conditions can't be fathomed legitimately. Typically these models include inert factors also to obscure parameters and recognized data perceptions. That is, either missing qualities exist among the data, or the model might be built up a great deal of only by forward the presence of extra in secret data focus. For instance, a blended model might be outlined a ton of only by forward that each decided data envelops a comparing in secret data, or inert variable, indicating the blend part to that each data has a place.

The Expected Maximization algorithmic standard result from the perception that there's the way to unwind these two sets of conditions numerically. One will only pick hasty qualities for one in all the 2 sets of questions, use them to appraise the subsequent set, at that point utilize these new qualities to look out an improved gauge of the essential set, at that point continue switching back and forth between the 2 till the resulting esteems each combine to attached focuses. It isn't clear this can work, anyway it is demonstrated that during this setting it will, which the side project of the probability is that (subjectively shut too) zero around then, that progressively means the reason for existing is either a most or a seat point. By and large, various maxima may happen, with not a single certification that the world most are going insight. A few probabilities even have singularities in them, i.e., strange maxima. For instance, in an extremely one in everything about arrangements that will be found by Expected Maximization in a blended model includes setting one in every one of the parts to have zero fluctuation and furthermore the mean parameter for a proportional component to be up to one in all the information focuses.

$$N(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{\sigma^2}} \dots (2)$$

There are numerous approaches to limit the degree of missing information. It might be useful to consolidate institutionalized guidelines to streamline information gathering, for example, preparing staff to gather and organize information accumulation, utilizing admirably characterized information definitions, and joining rationale and the limit checks for every datum component. Preliminary studies can distinguish factors, especially defenseless to missing qualities, and the process can be taken in order to improve its wholeness. [7] Regular observing of information quality and fulfillment gives basic criticism and specialists on the degree of data that is missing. [8] Furthermore, when gathering data about the personal satisfaction or other delicate issues, patients might be approached to give motivations to declining to take part, for example, an absence of time, issues getting language or long or too cozy surveys. This data can be utilized in the examinations of information and translation of the outcomes. Single imputation techniques don't represent the vulnerability of missing information, and therefore, standard blunders of the evaluations are probably going to be excessively little (in this manner overestimating the accuracy of the outcomes). This can potentially lead to Type I error (i.e., identifying an association when none exists). Mean attribution likewise does not safeguard the connections between factors; it just jams the mean of the watched information. In this way, if the information is MCAR, the gauge of the mean remains unbiased. [9] Under MCAR, if our point is to gauge implies (which is once in a while the principal focal point of research considers), mean imputation won't inclination the evaluations; it will just predisposition the standard blunders as referenced already. Since a large portion of the examination studies are keen on the connection among factors and not simply the mean, mean attribution ought to be evaded when all is said in done. It has been brought up already that last perception conveyed forward strategy can deliver one-sided evaluates in two bearings even under MCAR and have cautioned against utilizing this technique as the first or decision for dealing with missing data. [10] Multiple imputations take care of the issue of "excessively little or excessively huge" standard mistakes got utilizing conventional strategies for managing missing information. The point of multiple imputations is to give fair-minded and substantial evaluations of affiliations dependent on data from the accessible information, i.e., yielding assessments like those determined from full data. Missing information and consequently, different attribution may influence not just the coefficient gauges for factors with missing information but also the estimates for other variables with no missing data. Multiple Imputation is generally perceived as the standard strategy to manage missing information in numerous territories of research, and the technique has turned out to be progressively well known with the increasing accessibility of programming.

A full depiction of numerous attribution is past the extent of this paper. However, we give a concise review of its suppositions, execution, and methodologies.

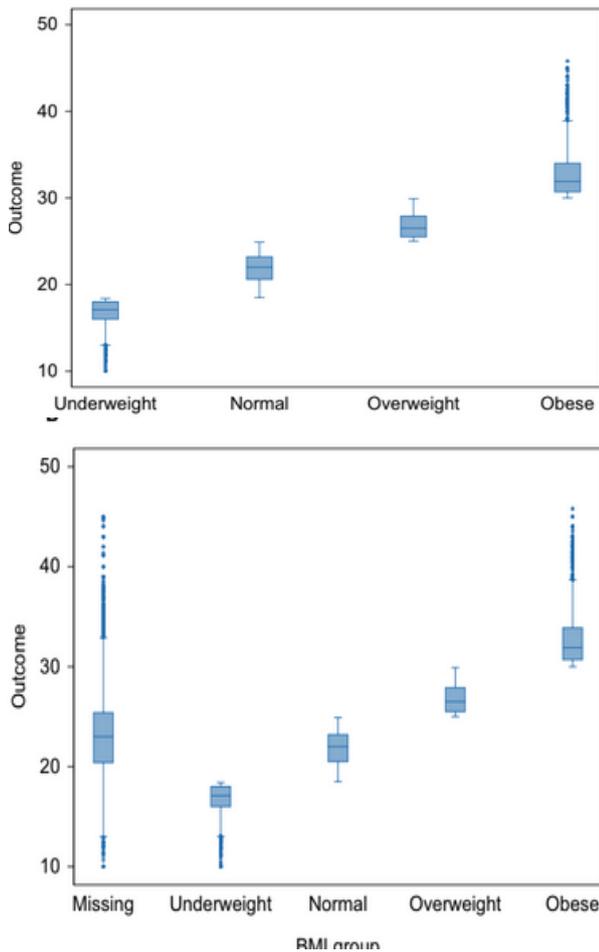


Fig. 1 Dispersion of BMI esteems by the result in full dataset (A) and in a dataset, with 35% missing qualities (B) for BMI has taken care of by making a missing BMI classification

Abbreviation: BMI - Body mass index

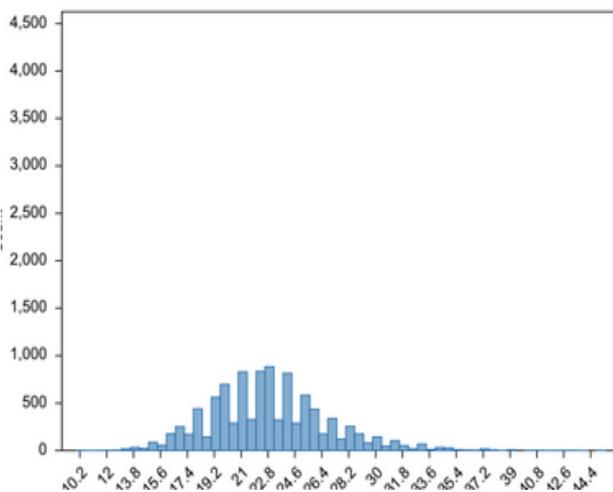


Fig. 2 Typical appropriation of watched BMI in a full dataset of 10,000 recordings

Abbreviation: BMI - Body mass index

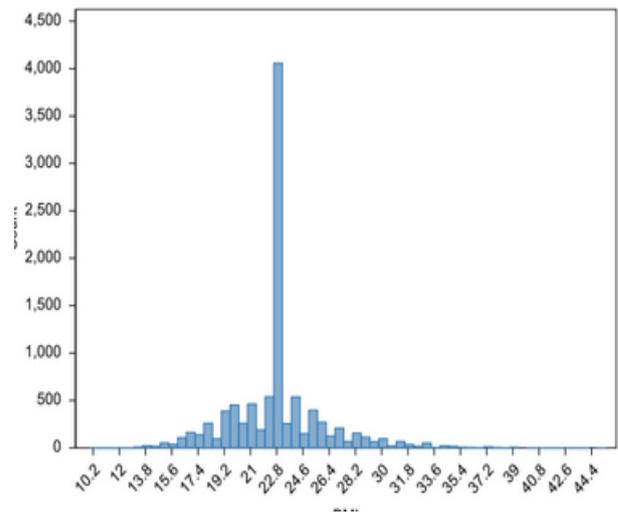


Fig. 3 Distribution of BMI in a dataset of 10,000 recordings, while 35% of BMI esteems is missing and supplanted by the watched mean BMI esteem

Abbreviation: BMI - Body mass index

The multiple imputation strategy in most measurable programming expands on the MAR assumption. However, the technique can deal with both MCAR and MNAR. Although we can't demonstrate whether the information is MAR, all things considered, in numerous situations, the MAR supposition that is increasingly conceivable when more factors are incorporated into the different attribution model.

Table. 1 An example of a situation when data are MAR rather than MCAR

Observed data	Patients with BMI value (%)	Patients with missing BMI value (%)
Smoking		
Yes	80	20
No	60	40
Comorbidity prior diagnosis		
Yes	85	15
No	25	75

Abbreviations: BMI, body mass index; MAR, missing at random; MCAR, missing completely at random.

The 3 NN algorithms were then applied to those frameworks:

Plain NN framework: the total set knowledge{of knowledge{of information} is employed per the hot-deck technique, and solely complete cases with no missing data C(X) are thought-about as donors.

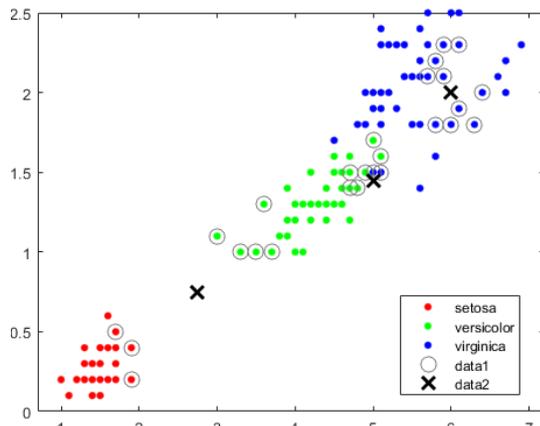


Fig. 4 k-NN Means Clustering example graph showing how different classes will be clustered

Filtered NN structure: Earlier imputation regards to the beneficiary X_i , the absolute set with zero missing learning $C(X)$ is sifted to pick through a lot of choices important to the missing set be amputated (X_{i_miss}). To the present completion, $C(X)$ is regarded as a wholesome set inside the setting of a relapse downside, any place the formula which has the missing data is known to be the ' X_{miss} ' is prepared in light of the fact that the classification variable and furthermore the elective letter factors (X_1, \dots, X_q) to be the indicators. Since, in actuality, things there's occasionally no sign on whether any connection exists among indicators and result or, if this connection exists, what kind it takes, an absolutely non-parametric decision algorithmic guideline is considered Associate in Nursing appropriate option. Inside the blessing setting, we tend to connect the RRelieFF algorithmic principle spoke to in; the set is then sifted to pick through $C_s(X) \subset C(X)$ any place ; $(X_1, \dots, X_s) \subset (X_1, \dots, X_q)$ with $s < q$. Inside the blessing setting. We also tend to set the measure of locals for RRelieFF up to ten and set s as ten nothing, twenty nothing or half-hour of the letter. The operation is performed separately for X_i as per each.

Banned NN structure: for each beneficiary X_i , from the arrangement of contributors with no missing information $C(X)$ with size m , an irregular arrangement of givers with size $m_n < m$ is picked with substitution like $CR(X) \subset C(X)$; the set $CR(X)$ is then acclimated attribute the missing worth for the beneficiary. The strategy is lasting n scope of times with totally extraordinary irregular portion of benefactors all together that numerous possible qualities $X_{iR1}, X_{iR2}, \dots, X_{iRn}$ from each $CR1(X), CR2(X), \dots, CRn(X)$ arbitrary portion; a definitively attributed value for X_i is determined on the grounds that the principal snapshot of the X_{iRn} credited qualities. In our exploratory arrangement, m_n is prepared to ten you take care of m and n to fifty arbitrary runs. The arbitrary technique of material, on condition that $mR/m \rightarrow 0$, is anticipated to be helpful in a very dataset with numerous uproarious factors that affect the investigation of the space from the beneficiary X_i to the contributors. As an aftereffect of clamor, cases that are disparate from X_i with the importance the properties correlative with the missing variable to be attributed (X_{i_miss}), or divisively picked as "close" to the beneficiary after they truly lie "far" from it. The material that givers with the previously mentioned compels, would encourage to

dispose of such givers and to push through the clamor produced by unessential alternatives remedying incompletely the excessively solid straightforwardness inclination inside the NN student because of NN is making mistaken presumptions concerning space and blunder is diminished by powerful it.

Random NN structure: for each beneficiary X_i an arbitrary arrangement of properties l like that $l = q^{\sqrt{}}$ is picked, and furthermore, the comparing set of benefactors with no missing information inside the l qualities $CL(X) \subset C(X)$ is then an idea about. The technique is lasting n scope of times with totally unique arbitrary portion of traits all together that numerous possible qualities $X_{iL1}, X_{iL2}, \dots, X_{iLn}$ are got from each $CL1(X), CL2(X), \dots, CLn(X)$ portion; a definitive imputed value for X_i is determined in light of the fact that the main snapshot of the X_{iLn} attributed values. In our trial arrangement n is prepared to fifty arbitrary runs. By and large, these structures contrast from c) in that randomization is presented inside the selection of properties rather inside the selection of benefactors. The enduring irregular selection of qualities would for each arbitrary run support the expulsion of incidental properties which will predisposition the space metric from the beneficiary to contributors and consequently cause problematic vicinities. The methodology is anticipated to have a high likelihood of progress once in a very dataset unessential (boisterous) characteristics number the non-uproarious properties correlative with the missing variable to be credited (X_{i_miss}) on the grounds that the chances are for keeping non-loud factors once the irregular decision of few qualities. All through each irregular run we tend to do hope to part the board the commotion all together that it's suited to a critical degree by the gathering in light of the fact that the "genuine" imputation esteems got from educational voters number the "arbitrary" or "false" attribution esteems got from non-instructive voters.

Full randomized NN framework: the methods represented in c) and d) are consolidated along in a very random forest-like fashion. In this way a double randomization is introduced within the learning algorithm: first a random set of $l = q^{\sqrt{}}$ attributes is chosen from the set of donors with no missing knowledge $C(X)$ in order that $CL(X) \subset C(X)$ so a random set of donors with size $m_n \ll m$ is chosen with replacement from $CL(X)$ to come up with a totally randomized set $CRL(X) \subset CL(X) \subset C(X)$. The procedure is repeated 50 times and, as in c) and d), the final imputed value is the same as the values derived in all the fifty random runs.

They are missing learning territory units generally characterized into the following three types: Missing completely at random (MCAR), Missing at random (MAR), and missing not at random (MNAR). At the point where individuals with missing information zone unit an arbitrary arrangement of the investigation populace, the possibility of being missing is that the equivalent for all cases; missing learning zone unit indicated as MCAR.

A case of MCAR is at one time a glass slide with analytic test material from a patient which was incidentally revealed indicated pathology and histology tests can not be performed, or once the individual has no weight estimated because the instrumentation was broken. Along these lines, under MCAR, missing information doesn't depend on either watched learning or in secret learning. In differentiation to MCAR, the term MAR is preposterous. Blemish happens once the missingness relies upon the information we have effectively found.

For instance, the information in a very discouragement overview is same to be MAR, given sexual orientation if men zone unit less conceivable than young ladies to round out the review. When sexual orientation is represented, the missingness doesn't depend on the degree of their downturn. Another case of MAR is once, in an investigation of weight, information on weight territory unit less conceivable to be recorded for more youthful individuals, because of they are doing not go to medicinal services offices as regularly as more seasoned individuals. At the point when the possibility that information zone unit missing relies upon the surreptitiously learning, similar to the value of the perception itself, at that point, the missing learning region unit indicated as MNAR.

For instance, overweight or scraggy individuals could likewise be a ton of conceivable to have their weight estimated than individuals with conventional weight, notwithstanding when age is represented. In this way, the clarification for the missingness is said to in secret qualities of the individual, and subsequently, learning zone unit MNAR. Another model is when individuals with serious despondency, or unfriendly impacts from prescription medicine, region unit a ton of or less conceivable to finish a study on discouragement. A third model is once information on monetary benefit zone unit missing, and along these lines, the opportunity of missingness is said to the degree of monetary profit, e.g., those with frightfully low or high monetary profit will not report their monetary profit.

For the principal half, in clinical medicinal strength examination, missing qualities region unit neither MCAR nor MNAR anyway MAR. Watched information will give U.S. some sign of whether missing learning territory unit MCAR, yet we will, in general, don't appear to be prepared, from this information alone or clear test, to pass judgment on whether missing information zone unit MAR or MNAR. By arranging the attributes of individuals with missing information against those while not, {we can|we will|we zone unit capable to} judge whether learning is probably going to miss obtaining on these qualities. We represent this in A model any place assortment of individuals zone unit lacking weight record (BMI) estimations.

In this model, we will see that among smokers, the extent of people with BMI found is higher contrasted with non-smokers. Likewise, among patients with recognized comorbidity earlier medical procedure, the extent of individuals with BMI found is higher looked at thereto of these while not recognized comorbidity. In this way, we will reason that the data don't appear to be MCAR. Diagram hypotheses are valuable in a very assortment of controls inside the fields of number juggling, building, software

engineering, and science to work out or judge the instruments of missingness.

In a therapeutic forte examination, causative graphical models, as coordinated non-cyclic charts (DAGs), can be utilized to affirm whether learning territory unit MAR, MNAR or MCAR, in this way illuminating the first worthy logical strategy to manage missing qualities.

IV. IMPLEMENTATION

Linear Regression

To start, a few indicators of the variable with missing qualities (for our situation 'Age' include) are distinguished utilizing a relationship grid. The best indicators are chosen and utilized as free factors in a regression equation. The variable with missing information is utilized as the objective variable.

The impediments of the model will, in general, exceed its point of interest. This shows that they underestimate standard blunders and, in this way, overestimate test insights. The primary rationalization is that the usurped qualities are totally controlled by a model connected to different factors, and they will in general fit together "excessively well," as it were, they contain no blunder. Likewise, one should expect that there is a straight connection between the factors utilized in the relapse condition when there may not be one.

k-Nearest Neighbor (kNN) Imputation

Despite the fact that XGBoost and Random Forest could likewise be utilized for information imputation, we will be examining KNN as it is broadly utilized. For k-Nearest Neighbor imputation, the missing data depend on a kNN algorithm. In this technique, k neighbors are picked dependent on some separation measure, and their normal is utilized as an imputation gauge.

$$(\sum_{ni=1}^k |x_i - y_i|)^{1/p} \dots (2)$$

The strategy requires the determination of the quantity of closest neighbors, and a separation metric. KNN can foresee both discrete (most regular incentive among the k closest neighbors) and persistent characteristics (mean among the k closest neighbors). The separate measurement shifts as per the kind of information:

Continues Data is the regularly utilized separation measurements for consistent information are Euclidean, Manhattan, and Cosine. Categorical Data uses hamming separation, where it is commonly utilized for this situation. It takes all the unmitigated traits and for each, tally one if the worth isn't the equivalent between two. The Hamming separation is then equivalent to the number of traits for which the worth was unique.

A standout amongst the most appealing highlights of the KNN is that it is easy to comprehend and simple to actualize. The principal inconvenience of utilizing kNN attribution is that it moves toward becoming tedious when breaking down enormous datasets.

Indiscriminant Expected Maximization Imputation Model using Multiple Classification Technique on Diabetic Dataset

Likewise, the precision of KNN can be seriously corrupted with high-dimensional information in light of the fact that there is little distinction between the closest and most distant neighbor. At last, the number of neighbors (k) must be painstakingly chosen when utilizing KNN imputation.

It isn't unexpected to have missing qualities in arrangement expectation issues if your successions have variable lengths. In the clinical research, it is necessary to avoid the missing data at maximum, and if you found how to deal with missing information in grouping expectation issues in Python with Keras.

In particular, we have learned through the experiment of how to code in order to evacuate columns that contain missing data. Masking layers to the missing data and power the model to become familiar with their importance. And importantly, how to effectively mask missing values to prohibit them from interfering with the calculations in the model.

V. EXPERIMENTAL RESULTS

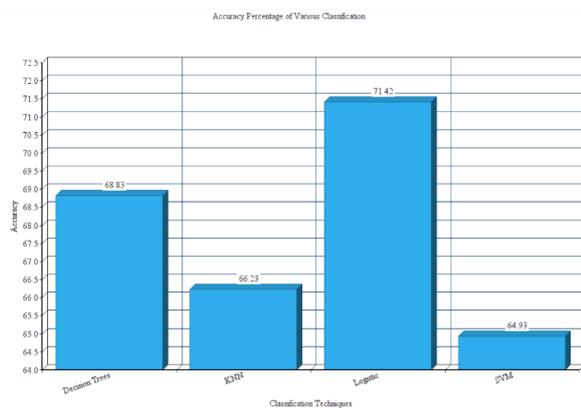


Fig. 5 Accuracy Percentage of Various Classification

A group of twenty, 003 biopsy instances of patients' present within the mortality statistics were derived from the Hull-Lifelab dataset. One category attribute and four input attributes (creatinine, sodium, urea, NT-proBNP) were enclosed in each instance. The missing worth dimension of every input attribute is listed in the tables. The distribution of the category attribute is illustrated in figure one.

The results of the experiments are given in Table 3, where the best classifier for every imputation technique is highlighted in bold. Once examining the results for every category worth, the predictions for death at 36m and 6m were further correct than the ones for 18m and 24m. This is in all likelihood due to the uneven frequency of the category attribute caused by a case-wise deletion in patients that didn't deliver the goods a minimum of 12, 18, 24 or 36m of follow-ups alone.

When we observed the results for every imputation technique, we found that the KNN classifier performs best for MCI, KMI, and FKMI; the choice tree classifier was the most effective for CMCI, KNNI, and SVM; the MLP classifier works best for EMI. Nevertheless, the Naïve Thomas Bayes classifier wasn't the most suitable classifier for any of the imputation ways. This could reveal that the trait independence assumption (of NB) wasn't true for our clinical HF data.

When we examined all the classifiers together; the CMCI method obtained the most efficient average exactness and recall. The EMI and SVM; ways had similar average precision and recall, and that they weren't removed from CMCI. The average exactness and recall obtained by four other imputation ways were much lower.

VI. FUTURE WORKS

In the general context of the clinical learning issue, the little size of the informational collection is the real obstacle. The limited sample size reduces our measurable power as far as possible and raises doubt about the appropriateness of our discoveries to the all-inclusive community. Cardiovascular attractive reverberation imaging is a moderately novel methodology in the assessment of pediatric cardiomyopathy and not part of the standard stir up, which confines the measure of information accessible for research. This absence of information propelled the investigation of attribution as a likely workaround.

With this heart dataset, the future work on imputation must center on master learning-is driven towards displaying missing attributes. Whilst remembering information is impossible because of the review idea of the dataset, attribution strategies could be executed explicitly on the grounds of social occasion, extra factors prone to be identified with the missing factors being referred to. This could appear as a similar variable estimated at a various time focuses or clinically corresponded factors with recently known connections to the missing variable. An attribution technique that depends on this turns out to be increasingly similar to extrapolation or interjection.

VII. CONCLUSION

We audited for missing price imputation and looked at the display of numerous imputation ways on a standard learning task. In this paper, we tried to establish those simple imputation ways (mean imputation), and a listwise cancellation was performed which is similar to a great deal of cutting edge imputation ways (choice trees, k-NN, and SOM). With our audits and test, we tend to infer that the imputation gives supportive information with regards to the toughness of a model learning philosophy, notwithstanding its effect on model execution. Combined with the view that simple imputation ways commonly perform just like a ton of cutting edge ones, we would support attributing missing attributes at whatever point conceivable.

REFERENCE

1. H. & S. Wang, "Mining the incomplete survey of data through classification," KnowlInfSyst, pp. 221, 2010.
2. Rubin DB, Inferenceandmissingdata,Biometrika.1976;63(3):581-592.
3. Graham JW, "Missing data analysis: making work in the real world," Annu Rev Psychol, 549-576;2009.
4. Donders AR: "A gentle intro to the imputation of missing values." J ClinEpidemiol.2006;59(10):1087, 1088, 1089, 1090 and 1091.
5. Mohan K & Pearl J, "Graphic models in inference to the missing data." Ed.Bottou L, Burges CJC, Welling M, Weinberger KQ & Ghahramani Z. Red Hook; 2013:1277-1285.

6. Alma B Pedersen, Tra My Pham, Pedersen, Irene Petersen, Nickolaj R Kristensen, "Missing and multiple data imputations in clinical, epidemiol research," 2017
7. Cappelleri JC, Symonds T, Zou KH, and Bushmakin A, "Patient-Reported Outcomes: Measurement, Interpretation & Implementation." Boca Raton, 2013.
8. Wisniewski SR, Leon AC, Otto MW, Trivedi MH, "Prevention in missing data for clinical research studies." 59(11):997-1000;2006.
9. Greenland S and Finkle WD. "Critics: A critical look at methods for handling missing covariates in epidemiologic regression." Am J Epidemiol. 142:1255-1265; 1995.
10. R. Deb and A. W.-C. Liew, "Missing value imputation for the analysis of incomplete traffic accident data," *Information sciences*, vol. 339, pp. 274-289, 2016.
11. C.-F. Tsai, *et al.*, "A class center based approach for missing value imputation," *Knowledge- Based Systems*, vol. 151, pp. 124-135, 2018.
12. M. Sangeetha, M. Senthil Kumaran, "Min Feature Replacement Algorithm Based Multiple Imputation Based Gap Analysis for Clinical Data." ISSN: 2278-3075, Volume-8 Issue-5 March, 2019.
13. M.D. Samad, *et al.*, "Predicting survival from large echocardiography and electronic health record datasets: optimization with machine learning," *JACC: Cardiovascular Imaging*, vol. 12, pp. 681-689, 2019.
14. A. B. Pedersen, *et al.*, "Missing data and multiple imputation in clinical epidemiological research," *Clinical epidemiology*, vol. 9, p. 157, 2017.
15. P. Rajasekar, M. Pushpalatha, "A Survey On Body Sensor Devices For Medical Services To Provide Miniature Design And Energy Harvesting Technique" *International Journal of Pure and Applied Mathematics* Volume 117 No. 16 2017, 87-92.