**Review 1 (Revision)**  
This work investigates four ML techniques when predicting diabetes. The author did a good job. I suggest to treat:  
1. Section 2. Please describe the PIMA dataset, as its description appears further, in section 3.1. ~~sa mut descrierea PIMA de la sect 3.2 la sect 2?~~   
2. Table 1. Please describe more how the 30,691 samples derived the 445,132 samples from the last column. (solved -added detalies in the last paragraph from Literature review)  
Why the feature age was not preserved as an important feature? ~~Nu am neaparat o motivatie pentru ca pur si simplu am omis. As putea zice ca m-am focusat mai mult pe aspecte ce tin de partea medicala/de sanatate. Dar e necear sa include asta?~~  
3. It is unusual to mention Table 3 at page 7, and to put it at page 12 ~~Sa mut tabelul la pag 7? Nu are mai mult sens sa ramana la finalul tuturor experimentelor?~~4. The numerical experiment needs description of the computing environment, architecture, apps and language. (solved -added subsection Environment on page 6)  
5. It is not clear why some samples were deleted (page 10, 424,045 + 20,000 differs from the value in Table 1. (solved – mismatch in table 1)  
   
  
**Review 2 (Acceptance)**  
The paper performs a comparative analysis of machine learning models for predictind the risk of diabetes: it looks at RFC, Naive Bayes, DTC, and LightGBM. Along the way it also constructs a new dataset by filtering and processing data from an extensive medical survey : BRFSS.. It also employs a smaller dataset, PIMA. The paper is very well articulated: it contains clearly formulated research questions, an extensive survey of existing work and several experiments which are well-designed and shed light on the originally formulated research questions. Besides comparing the models per se it also performs an analysis of feature importance (where pertinent).  
  
Detailed comments, typos, suggestions:   
  
section 1,2: Section 1,2 (capital s)  
data set -> dataset (in several places)  
but removing -> removed first  
limited on a short period-> to a short period  
on the just obtained dataset -> skip  
educations ->education  
Table 1 is somehow confusing: its caption says that it contains a comparision of datasets and results between literatuure studies and the current one; at the same time it contains only some of the newly obtained results; the results concerning the PIMA dataset are not mentioned in the table. It is also not clearly demarked what is new and what previous work. (solved-added another column for PIMA results for my experiments)

to be not -> to not be  
defending depressive\_disorder-> I think it i meant defeating  
missing reference ??  
difficulty\_walking, walking\_difficulty  
What it brings over SMOTE -> It adds to SMOTE a random value  
so instead -> so that instead  
see Figure 4 ) ->extra space  
Despite this improvement appearance -> apparent improvement  
SelectKBest was used in two experiments [...] the paragraph is somehow redundant; it has asll been explained earlier  
feature importance analysis carried-> was carried out  
  
**Review 3 (Acceptance)**  
The paper presents a solid comparative analysis of several ML algorithms -  Random Forest Classifier (RFC), LightGBM, Decision Tree (DTC) and Naive Bayes - for the prediction of diabetes. Experiments are well designed and results are presented for two datasets. Feature importance analysis is also addressed.  
The experimental section can be improved by adding a final subsection 'Discussion' to present some final remarks on the experimental results and the answears to the research questions defined in the Introduction section. (solved- added discussion subsection)