Anchor point - conversation Agent
The following are my notebooks -
1.Train_reader_edit- It may look same as mandate 3 but I have done hyperparameter tuning of the model on my dataset (link) . After 10 trials each with 9 epochs in random_search and nearly 40 epochs in model.fit the results were as follows compared to the previous one-
Exact match Score(now)- .27
Exact match Score (then) - .17 to .23
A whopping 4% increase :)

I have saved all the models and uploaded them on Kaggle.
2.Test_reader- A very short notebook in which I Tried asking the question giving the context on the model saved previously.
(Feel free to import any model and check how well it answers the question, Suprisingling the model trained on Squad performs better than the model trained on my dataset)

3.information_retrival- Again a short notebook that feeds the document/ database to BM25 for information retrieval task. I have done some preprocessing and text cleansing before feeding the contexts to BM25.

4.reader_retriver- If you are in a hurry please go right to this as it combines everything from all previous notebooks' end results you only have to provide the document and the question and the scripts return the answer and the context, the retriever finds the context in the document relevant to the question and the reader model returns the answer to the query from the context given by retriever.
All the models, datasets, and documents are available here.

Scope Of improvement-
I am sure there are many but to list a few-
 1. There must have been a better way to create the dataset rather than manually annotating it. In fact, the dataset itself could have been better.
2. there must have been a better way to finetune and parameter tuning
3. There are many information retrievers out there some even have streaming and indexing, so this could have been improved as well
4. There could be more than one way to ask questions so preprocessing the question in a better way is also a possible improvement

Please feel free to reach out to me in case you have any suggestions or wanna point out any mistakes, any criticism with good intent is much appreciated.

References-
https://www.analyticsvidhya.com/blog/2021/05/build-your-own-nlp-based-search-engine-using-bm25/
https://qa.fastforwardlabs.com/
https://github.com/google-research/bert