Broad Functionality: **Real-time prediction & Batch prediction**

Principle: **Deployment and Prediction**

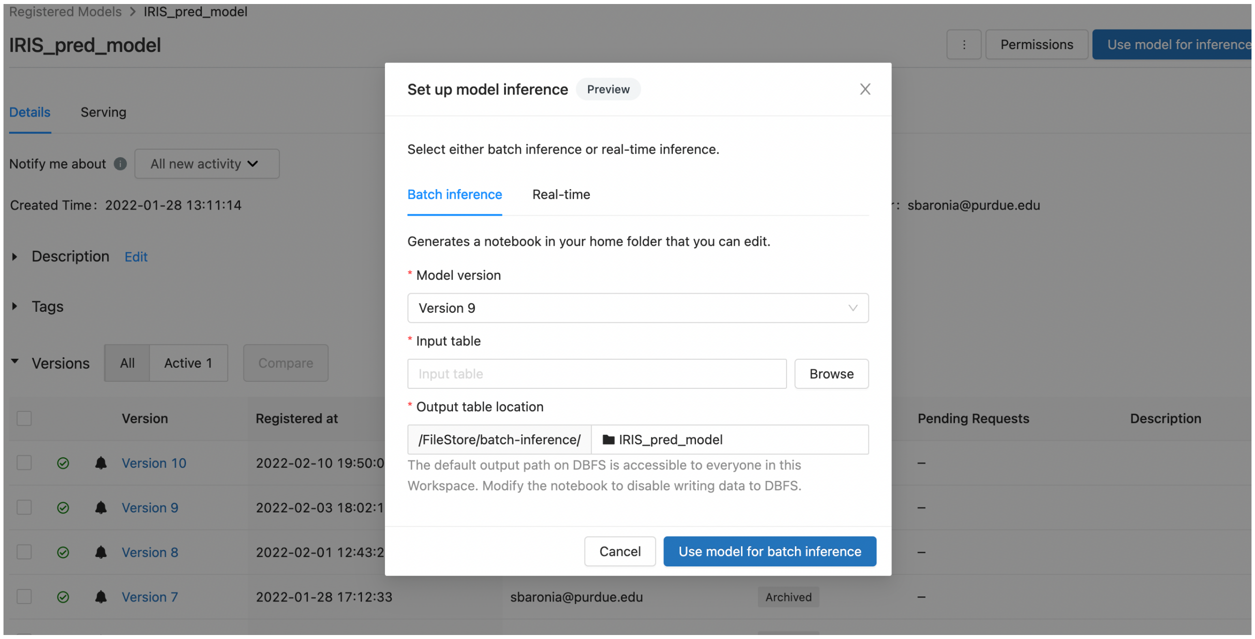
Understanding Functionality: Real-time predictions are also known as on-demand predictions, that is Predictions are generated in real-time using the input data that is available at the time of the request, the process must be completed in a fraction of seconds, to avoid latency in the generation of the result. And Batch predictions are a method of making predictions with large volume datasets, in which you pass input data and get predictions for each of your records. And as an output, we get predictions files.

Does MLflow supports the functionality *directly*: **Yes**

- Explanation: MLflow helps to generate code for batch or real-time inference. Leveraging the MLflow module Model Registry, we can automatically generate a notebook using the large dataset. In the MLflow Run page for your model, we can copy the generated code snippet for inference on pandas or Apache Spark DataFrames. we can also customize the code generated by either of the above options.
- Extra remark: NA

Does MLflow supports the functionality *indirectly*: **NA**

Screenshots for MLflow supports the functionality *directly or indirectly*:



Select either batch inference or real-time inference.

Generates a notebook in your home folder that you can edit.

* Model version: Version 9

* Input table: Input table

* Output table location: /FileStore/batch-inference/ IRIS_pred_model

The default output path on DBFS is accessible to everyone in this Workspace. Modify the notebook to disable writing data to DBFS.