

Summary

- When working on deploying a model, we need to check if the feature engineering pipeline is the same in the model development code and inference code.
- Developers sometimes develop a model in an environment (e.g. on laptop) other than the deployment environment. When python package versions in these two environments are different, feature engineering pipeline mismatch happens.
- The following diagram focuses on pipeline mismatches originated from package version differences, although mismatches can also due to other reasons.

Vocabularies

^ Dev and Deploy: the development environment where model development is done, and the deploy environment where model inference is done.

* Modelling Decisions: decisions on feature engineering, hyper-parameter set, model/feature stability, classification decision threshold, etc.

† Feature Engineering Pipeline Matching: compare the inference scores generated these two ways

Package version used in making *	Package version in deployment environment
Modelling Decisions	
Modelling feature engineering code	Deployment feature engineering code
Same testing dataset	
Same trained model	
Inference scores	Inference scores



Diagram

