**Feature Finder In Pega**

**Feature Finder is a tool designed to assist data scientists working with CDH in identifying the most suitable predictors for their ADM models.**

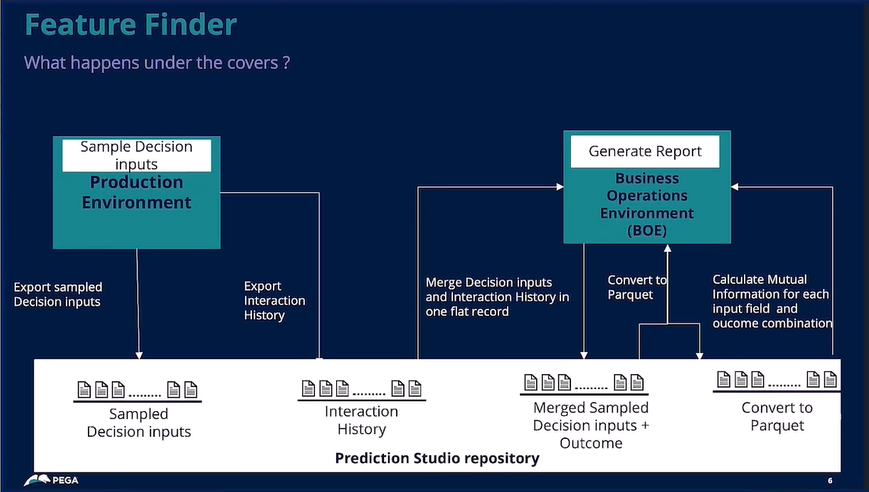
**In a typical CDH environment we temporarily cache customer profile data or xCAR to derive insights cache(CIC)**

**So this CIC contains lot of features, often upto 4,000.**

**While many of these features can be utilized as predictors, its often not possible.**

**That is because ADM model has a service limit of 500 to 700 so that number of predictors can be handled since it is a real time model.**

**So basically due to this limitation, data scientists currently resort to a manual process where they select predictors based on their intuition, their experience. However approach is time consuming may not always be the most optimal result because there is no data science behind it.**

****

**To address this challenge, the feature finder scans the CIC and suggests the top performing predictors for each ADM model on a bi-weekly basis to help the data scientists.**

**Feature finder operates by creating a comprehensive data layer tht will combine CIC with Interaction histories for sampled inbound interactions**

**The process can be enabled by simple set of configurations in the production and BOE environment in the prediction studio. This just enables the data migration pipeline.**

**The process usually works in 3 steps**

* **First in production we sample the data is exported from the production environment on a daily basis and stored into prediction studio repository. This repository is shared along with the BOE environment.**
* **Secondly data is imported from this repository back into BOE environment because it is shared and response each of the sampled decision is stitched together with interaction history of the data to create this comprehensive data layer where we have decisions being made and the response that are coming in from interaction history and all tied together.**
* **Finally we came to analysis process. And crucially this also runs in the BOE environment, ensuring that that does not impact the production. This allows data scientists to use metrics like mutual information. Where we go along and explore the predictors for each and every ADM model.**

**Moreover the data set generated by Feature finder holds immense values and data scientists have told us as such that this decision versus response with mutual metric information that you’re sort of creating up can be used for different analysis beyond the ADM Feature Finder or beyond the ADM model selection process also.**

**Feature finder ultimately generates a detailed report on bi-weekly basis. And this timeline is configurable and this report provides users with valuable insights and recommendations regarding the potential best predictors for each and every ADM model.**

**And essentially this empowers data scientists to make real data science driven decisions for their model development process.**

**So we combine customer insights cache (CIC) and interaction history and basically we review all our past performance in real time. See what predictors have worked best for us. And we create a rank list to suggest which are our best predictors. Again, for each and every model. And we create a rank list to suggest which are our best predictors.**

**And all this is done without impacting our current models or our production environment.**