Work Plan – Appendix

Goal: Solution for chemical fingerprinting used to represent molecules for property prediction in downstream applications.

Assumptions:

1. BPC has domain experts who can help with feature engineering for machine learning model.
2. BPC has a quick turnaround time for security review for new infrastructure to be provisioned for AI/ML workloads.
3. BPC has employees who can also be part of the project and work with the consultants to help source data from upstream systems.
4. Downstream systems have the necessary budget to work with the project to develop services to consume API endpoints.
5. Downstream systems can provide continuous feedback or provide access to their datastores to improve the model on a regular basis.
6. BPC did some POC before deciding on the algorithm ChemBERTa.
7. BPC has data that will meet the threshold to get better prediction and can be sourced in both batch and streaming format.
8. If enough training data is not available, then use data from PubChem.
9. Some of the components mentioned in solution design is Google Cloud specific, but if the client is already component in other cloud providers easier cloud native alternatives can be used.
10. The proposed design for deployment and exposing the model is though containerization and using Kubernetes. If the client prefers a different strategy we can work with them to finalize on the right fit.

Clarifying questions:

1. How long is the onboarding process, for now onboarding time isn’t included in project timeline.
2. Enterprise architecture team will be available to get the necessary approval for new technologies to be used to support the project.
3. Does the client have a data science team today who has done something similar.
4. How will be operations/support for this product look like once its deployed to production.
5. How quickly can be downstream systems provide feedback after using the model prediction?
6. How is change management handled as this will be a mindset change for the client.
7. What is the existing mechanism to track risks associated with the project and who is the in the line of escalation.