AI for Data Scientist

1. Vertex AI:

unified machine learning platform will help you build, deploy and scale more effective AI models.

Benefits:

- Accelerate ML with a unified data and AI platform and tooling for pre-trained and custom models
- Build generative AI apps quickly and responsibly with <u>Model Garden</u> and <u>Generative AI Studio</u>
- Implement MLOps practices to efficiently scale, manage, monitor, and govern your ML workloads
- Reduce training time and costs with <u>optimized infrastructure.</u>

Key Features:

- Choose a model that fits your needs (using Model Garden and Generative AI studio)
- End-to-end MLOps (train, test, monitor, deploy, and govern ML models at scale)
- Low-code and no-code tooling
- Open and flexible AI infrastructure

All features:

- **Vertex AI Model Garden:** A single place to search, discover, and interact with a wide variety of foundation models from Google and Google partners, available on Vertex AI
- **Vertex AI Generative AI Studio:** A managed environment in Vertex AI that makes it easy to interact with, tune, and deploy foundation models to production
- **AutoML:** Easily develop high-quality custom machine learning models without writing training routines. Powered by Google's state-of-the-art transfer learning and hyperparameter search technology.

- **Deep Learning VM Images:** Instantiate a VM image containing the most popular AI frameworks on a Compute Engine instance without worrying about software compatibility.
- **Vertex AI Workbench:** Vertex AI Workbench is the single environment for data scientists to complete all of their ML work, from experimentation, to deployment, to managing and monitoring models. It is a Jupyter-based fully managed, scalable, enterprise-ready compute infrastructure with security controls and user management capabilities.
- **Vertex AI Matching Engine:** Massively scalable, low latency, and cost-efficient vector similarity matching service.
- **Vertex AI Data Labeling:** Get highly accurate labels from human labelers for better machine learning models.
- **Vertex AI Deep Learning Containers:** Quickly build and deploy models in a portable and consistent environment for all your AI applications.
- **Vertex AI Feature Store:** A fully managed rich feature repository for serving, sharing, and reusing ML features.
- **Vertex ML Metadata:** Artifact, lineage, and execution tracking for ML workflows, with an easy-to-use Python SDK.
- **Vertex AI Model Monitoring:** Automated alerts for data drift, concept drift, or other model performance incidents which may require supervision.
- **Vertex AI Neural Architecture Search:** Build new model architectures targeting application-specific needs and optimize your existing model architectures for latency, memory, and power with this automated service powered by Google's leading AI research.
- **Vertex AI Pipelines:** Build pipelines using TensorFlow Extended and Kubeflow Pipelines, and leverage Google Cloud's managed services to execute scalably and pay per use. Streamline your MLOps with detailed metadata tracking, continuous modeling, and triggered model retraining.
- **Vertex AI Prediction:** Deploy models into production more easily with online serving via HTTP or batch prediction for bulk scoring. Vertex AI

Prediction offers a unified framework to deploy custom models trained in TensorFlow, scikit or XGB, as well as BigQuery ML and AutoML models, and on a broad range of machine types and GPUs.

- **Vertex AI Tensorboard:** This visualization and tracking tool for ML experimentation includes model graphs which display images, text, and audio data.
- **Vertex AI Training:** Vertex AI Training provides a set of pre-built algorithms and allows users to bring their custom code to train models. A fully managed training service for users needing greater flexibility and ==customization or for users running training on-premises or another cloud environment.==
- **Vertex AI Vizier:** Optimized hyperparameters for maximum predictive accuracy.

2. **Vertex AI Workbench:**

The single development environment for the entire data science workflow(data analytics and machine learning workflows).

AI for Developer

AI infrastructure