What real-time data can we use?

Realtime, the essential data that is needed for predictions would be

* FromISO,
* ToISO,
* FromZipCode,
* ToZipCode,
* ForwardingAgentName,
* CO2,
* MeansOfTransport,
* TotalTransportCost,
* NumberOfPieces,
* Distances,
* Weights, and
* Agents.

What infrastructure is needed?

To run the models for prediction, a minimalist backend server should have

* 4-8GB of RAM,
* 5-10GB of Storage space as a HDD or an SSD and
* a CPU that has atleast 2 cores and with a clock-speed of 2.0GHz.

This is largely due to the saved models of Random Forest being as large as 700-800Mb in size.

Language model?

The best model for our purpose, after trying out multiple models, turns out to be the Random Forest model, only slightly better than the Gradient Boosted Trees Model.

Although regression models are predictive models, they cannot be considered a Language model as they serve different purposes and are applied to distinct contexts.