A. PREPARE TRAINING DATA.

This piece focuses on collecting real reviews about gig workers. It currently functions as a web crawler that collects data from websites like SiteJabber, and ConsumerAffairs that share real-world reviews about gig workers. Once the data is collected, the module focuses on labeling each review based on whether it focuses on mission-critical metrics or not (factors that the worker controlled or not) The labeling is done by analyzing the policies of each gig market and is currently completed by crowd workers.

Notice that we worked with three of the most popular gig markets. %in their field.
For each of these three gig markets, we trained our tool to detect reviews that involved factors outside a worker's control. For this purpose, for each gig market we: (1) collected 1,000 real-world reviews from SiteJabber (for the three scenarios); (2) had two independent college graduate coders classify each of these reviews into whether they involved worker's performance or factors outside the worker's control. We provided summaries of what factors were considered to be within the worker's control and examples of which ones were not. Coders were also given a link to the policies of each of the three gig markets to better assess the variables that the marketplace considers are under a worker's control. Some explained examples were given to coders to have a common agreement when dealing with ambiguous cases. The two coders agreed on the classification of 94.7\% of all the reviews (Cohen's kappa =.86: Strong agreement). We then asked a third college graduate coder to act as a tiebreaker in cases of disagreement. After this step, for all three types of gig markets, we had a labeled set of reviews. The labeled data was provided as input to Reputation Agent's Smart Validator to train its models.

TRAIN AND TEST INTELLIGENT MODEL.

Reputation Agent uses stratified sampling to split the labeled data into training, test, and validation sets under proportions of 80%, 10%, and 10% (the validation set helps to avoid overfitting). Using Python 3 and the Keras framework with Tensorflow, we trained a model to learn to recognize reviews that evaluate workers based on mission-critical metrics and non-critical ones. We train the following models:

Word ngram + LR: Logistic regression with word ngrams.

Char ngram + LR: Logistic regression with character ngrams.

(Word + Char ngram) + LR: Logistic regression with word and character ngrams.

RNN no embedding: Recurrent neural network (bidirectional GRU) without pre-trained embeddings.

RNN + GloVe embedding: Recurrent neural network (bidirectional GRU) with GloVe pre-trained embeddings.

CNN (multi-channel): Multi-channel Convolutional Neural Network.

RNN + CNN: Recurrent neural network (Bidirectional GRU) + Convolutional Neural Network.

Google BERT (Devlin et al., 2018): Bidirectional Encoder Representations from Transformers, is a new method of pre-training language representations which obtains state-of-the-art results on a wide range of Natural Language Processing (NLP) tasks.

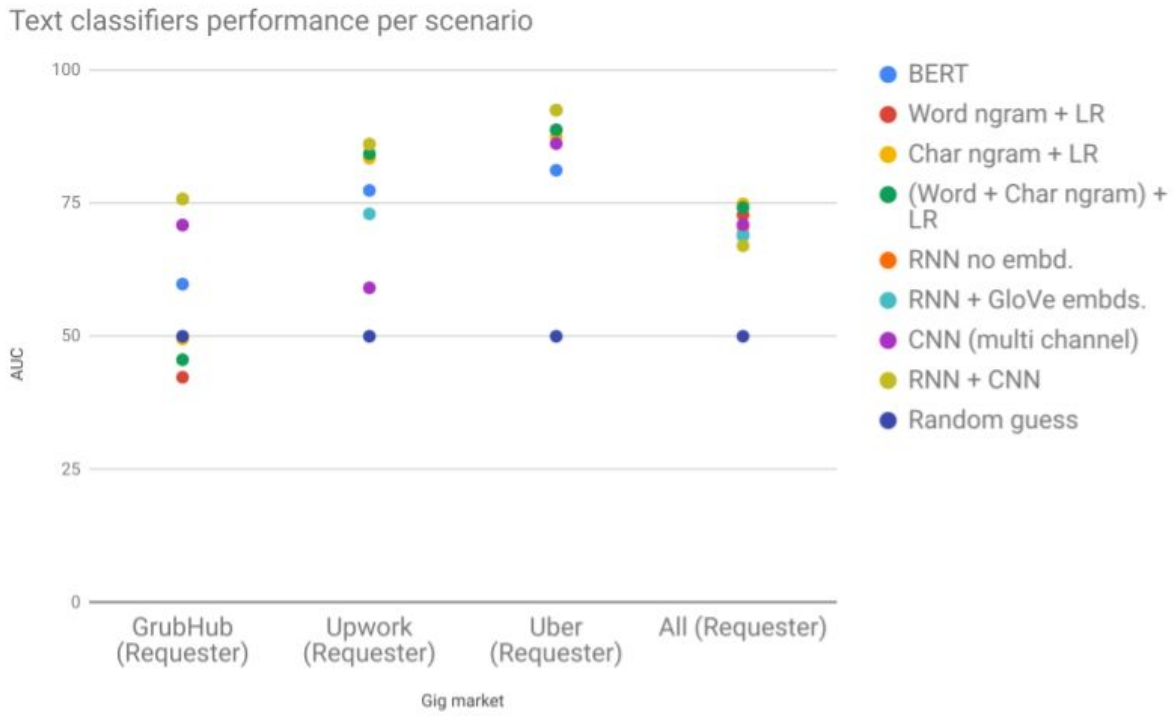

Figure X. Text classifier benchmark. Recurrent Neural Networks (RNN) + Convolutional Neural Network (CNN) approach performed better across conditions.

We implemented early stopping as a method to stop training once the model performance stops improving on a hold out validation dataset. For the deep learning models, we used a binary cross-entropy loss function, ADAM as an optimizer, and a learning rate of 0.001.

Fig. X presents an overview of the benchmark of Reputation Agent's training models. We note that different machine learning models performed better on different gig markets. However, RNN (Recursive Neural Network, a Deep Learning Algorithm) performed in general best across all gig markets. Once Reputation Agent had identified the best models to use for a particular gig

market on the trained condition, then it is ready to be deployed on the gig market. After the model has been trained, it is exposed as a REST web service via JSON requests to our front-end interface. The service is consumed directly by Reputation Agent’s Accuracy Promoter and it displays the messages accordingly.