The algorithm I implemented works by (1) preprocessing headlines using basic text manipulation, (2) extracting word tokens contained in headlines, (3) transforming tokens into feature vectors using pre-trained word embeddings, and finally (4) using these features vectors as the input of a regression algorithm following a supervised training paradigm.

1. Preprocessing

- Removal of punctuation
- Correcting for capitalization
- Removal of hyphens designating word to be replaced by edit word

2. Tokenization

Tokenization performed primarily using the SpaCy Python library:

- Removal of "stop word" (words that common and carry little information, e.g. "the", "to")
- Splitting remaining text of headlines into their words (i.e. tokens), preserving only those words that are represented in the pre-trained word2vec model (GoogleNews, 300-dim.).
- Combining the tokens composing named entities into a single token (e.g. "New" + "York" becomes "New York")

3. Feature Vectors

- Transformation of each token into its word2vec vector
- Concatenation of (1) average vector of headline tokens, (2) replaced word vector, and (3) edit word vector
- The resultant vector is our feature vector

4. Regression

I experimented with four different algorithms to serve the role of this latter regression algorithm (all available in the scikit-learn Python library):

(1) Simple linear regression (sklearn.linear_model.LinearRegression)
(2) Lasso regression (sklearn.linear_model.Lasso)
(3) Ridge regression (sklearn.linear_model.Ridge)
(4) Multilayer perceptron regression (sklearn.neural_network.MLPRegressor)

In order to determine the best hyperparameters for each regressor so as to better determine which of the four would serve as the best regression algorithm, I used grid search to exhaustively search for the combination of hyperparameters that resulted in the smallest validation error. I believe that the respective pools of possible hyperparameters I decided on were reasonable in breadth.

Regression Algorithm	**Validation Error (RMSE)**	**Test Error (RMSE)**
Simple linear regression	0.24446	0.55168
Lasso regression	0.34373	0.57232
Ridge regression	0.26516	0.53008
Multilayer perceptron regression	0.24401	0.53389

The best algorithm was found to be ridge regression. Its best hyperparameters were:

- alpha: 1.0
- normalize: True
- tol: 0.001