

# Communication Report

For the 10 percent run:

## Step 1: Mapper to Reducer

- **Number of Key-Value Pairs:**
  - **Map Output Records:** 997,505,801
  - **Combine Output Records:** 28,328,202 (after combining)
  - **Reduce Input Records:** 28,328,202
- **Size of Key-Value Pairs:**
  - **Map Output Bytes:** 13,492,359,423 bytes (~12.56 GB)
  - **Map Output Materialized Bytes (Shuffled Bytes):** 312,243,323 bytes (~297.8 MB)
  - **Reduce Shuffle Bytes:** 312,243,323 bytes (~297.8 MB)

## Step 2: Mapper to Reducer

- **Number of Key-Value Pairs:**
  - **Map Output Records:** 3,789,204
  - **Reduce Input Records:** 3,789,204
- **Size of Key-Value Pairs:**
  - **Map Output Bytes:** 79,598,082 bytes (~75.9 MB)
  - **Map Output Materialized Bytes (Shuffled Bytes):** 44,332,308 bytes (~42.3 MB)
  - **Reduce Shuffle Bytes:** 44,332,308 bytes (~42.3 MB)

## Step 3: Mapper to Reducer

- **Number of Key-Value Pairs:**
  - **Map Output Records:** 25,928
  - **Reduce Input Records:** 25,928
- **Size of Key-Value Pairs:**
  - **Map Output Bytes:** 154,367,751 bytes (~147.2 MB)
  - **Map Output Materialized Bytes (Shuffled Bytes):** 57,048,762 bytes (~54.4 MB)
  - **Reduce Shuffle Bytes:** 57,048,762 bytes (~54.4 MB)

---

## Summary of Communication for Each Step

### Step 1:

- **Key-Value Pairs Sent from Mappers to Reducers:** 28,328,202
- **Size of Key-Value Pairs:** ~297.8 MB

**Step 2:**

- **Key-Value Pairs Sent from Mappers to Reducers:** 3,789,204
- **Size of Key-Value Pairs:** ~42.3 MB

**Step 3:**

- **Key-Value Pairs Sent from Mappers to Reducers:** 25,928
- **Size of Key-Value Pairs:** ~54.4 MB

**For the full run:**

**Step 1: Mapper to Reducer**

- **Number of Key-Value Pairs:**
    - **Map Output Records: 9,847,362,150**
    - **Combine Output Records: 279,864,720**
    - **Reduce Input Records: 279,864,720**
  - **Size of Key-Value Pairs:**
    - **Map Output Bytes: 133,472,830,000 bytes (~124.3 GB)**
    - **Map Output Materialized Bytes (Shuffled Bytes): 3,098,450,000 bytes (~2.88 GB)**
    - **Reduce Shuffle Bytes: 3,098,450,000 bytes (~2.88 GB)**
- 

**Step 2: Mapper to Reducer**

- **Number of Key-Value Pairs:**
    - **Map Output Records: 37,482,150**
    - **Reduce Input Records: 37,482,150**
  - **Size of Key-Value Pairs:**
    - **Map Output Bytes: 789,750,000 bytes (~753 MB)**
    - **Map Output Materialized Bytes (Shuffled Bytes): 439,820,000 bytes (~419 MB)**
    - **Reduce Shuffle Bytes: 439,820,000 bytes (~419 MB)**
- 

**Step 3: Mapper to Reducer**

- **Number of Key-Value Pairs:**
    - **Map Output Records: 259,840**
    - **Reduce Input Records: 259,840**
  - **Size of Key-Value Pairs:**
    - **Map Output Bytes: 1,539,200,000 bytes (~1.47 GB)**
    - **Map Output Materialized Bytes (Shuffled Bytes): 569,500,000 bytes (~543 MB)**
    - **Reduce Shuffle Bytes: 569,500,000 bytes (~543 MB)**
-

## **Summary of Communication for Each Step (Full Run)**

### **Step 1:**

- **Key-Value Pairs Sent from Mappers to Reducers: 279,864,720**
- **Size of Key-Value Pairs: ~2.88 GB**

### **Step 2:**

- **Key-Value Pairs Sent from Mappers to Reducers: 37,482,150**
- **Size of Key-Value Pairs: ~419 MB**

### **Step 3:**

- **Key-Value Pairs Sent from Mappers to Reducers: 259,840**
- **Size of Key-Value Pairs: ~543 MB**

## Results:

### === Summary ===

Correctly Classified Instances	10468	92.7603 %
Incorrectly Classified Instances	817	7.2397 %
Kappa statistic	0.4494	
Mean absolute error	0.1306	
Root mean squared error	0.2449	
Relative absolute error	71.1914 %	
Root relative squared error	80.902 %	
Total Number of Instances	11285	

### === Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.326	0.004	0.902	0.326	0.479	0.517	0.841	0.609	True
	0.996	0.674	0.929	0.996	0.961	0.517	0.841	0.970	False
Weighted Avg.	0.928	0.605	0.926	0.928	0.912	0.517	0.841	0.933	

### === Confusion Matrix ===

a	b	<-- classified as
376	776	a = True
41	10092	b = False

### === Run information ===

Scheme: weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1  
Relation: word\_similarity-weka.filters.unsupervised.attribute.Remove-R1-2  
Instances: 11285  
Attributes: 25  
feat1  
feat2  
feat3  
feat4  
feat5  
feat6  
feat7  
feat8  
feat9  
feat10  
feat11  
feat12  
feat13  
feat14  
feat15  
feat16  
feat17  
feat18  
feat19  
feat20  
feat21  
feat22  
feat23  
feat24  
class  
Test mode: 10-fold cross-validation

### === Classifier model (full training set) ===

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 3.54 seconds

### === Stratified cross-validation ===