

# FastText Model Comparison Report

## 1. Introduction

**This report compares the results from three stages of training a FastText model:**

1. Custom FastText Model trained on Yelp data.
2. Pretrained FastText Model (Facebook's cc.en.300.bin).
3. Updated Pretrained Model with Yelp data.

We analyze the top 10 similar and opposite words for two selected words: "service" and "work" across all three models.

## 2. Model Comparison

### 2.1 Custom FastText Model (Trained on Yelp Data)

Top 10 Similar Words to "service":

1. servicea (0.9383)
2. serviceand (0.9202)
3. servicevery (0.9123)
4. serviceno (0.9089)
5. servicei (0.9070)
6. servicejust (0.9040)
7. servicethe (0.9024)
8. servicestaff (0.8998)
9. servicequality (0.8969)
10. servicegood (0.8964)

Top 10 Opposite Words to "service":

1. magic (0.0121)
2. maze (0.0051)
3. pirate (0.0015)

4. field (0.0011)
5. battle (-0.0002)
6. phrase (-0.0006)
7. packet (-0.0144)
8. wiggle (-0.0198)
9. bunny (-0.0207)
10. heart (-0.0251)

Top 10 Similar Words to "work":

1. workplace (0.7484)
2. working (0.7318)
3. hardworking (0.7231)
4. worked (0.7100)
5. workweek (0.6993)
6. workman (0.6853)
7. teamwork (0.6815)
8. communicator (0.6418)
9. workin (0.6400)
10. workmanship (0.6368)

Top 10 Opposite Words to "work":

1. grape (0.0368)
2. dolmas (0.0354)
3. gran (0.0333)
4. goose (0.0333)
5. trio (0.0298)
6. dolma (0.0194)
7. largest (0.0144)

8. heaven (0.0139)
9. habanero (0.0106)
10. nugget (0.0026)

## **2.2 Pretrained FastText Model (Facebook's cc.en.300.bin)**

Top 10 Similar Words to "service":

1. servicee (0.7351)
2. service (0.7327)
3. sevice (0.7232)
4. serve (0.7060)
5. services (0.7012)
6. service.The (0.6951)
7. service.This (0.6788)
8. servcie (0.6641)
9. service (0.6479)
10. service.It (0.6473)

Top 10 Opposite Words to "service":

1. Contorted (0.2277)
2. AbiNormal (0.2057)
3. teensie (0.2056)
4. Mythomania (0.2052)
5. Squinch (0.2049)
6. R1- (0.2022)
7. Ghostfreak (0.2011)
8. Sixo (0.2005)
9. 825g (0.2000)
10. 1Even (0.1999)

Top 10 Similar Words to "work":

1. working (0.7022)
2. works (0.6820)
3. work.But (0.5919)
4. Work (0.5831)
5. work.The (0.5739)
6. job (0.5733)
7. worked (0.5726)
8. work.I (0.5692)
9. work.It (0.5630)
10. work.However (0.5610)

Top 10 Opposite Words to "work":

1. 20Table (0.2364)
2. 102012 (0.2301)
3. 28Table (0.2296)
4. RGLS (0.2288)
5. 0.639 (0.2283)
6. Record-Low (0.2274)
7. 49869 (0.2267)
8. 30Table (0.2260)
9. 4Section (0.2241)
10. 293.0 (0.2238)

## 2.3 Updated Pretrained FastText Model (After Training on Yelp Data)

Top 10 Similar Words to "service":

1. servicesAdvice (0.9905)
2. services (0.9889)
3. service--and (0.9887)
4. service.And (0.9884)
5. service.Now (0.9883)
6. service.Also (0.9882)
7. service.Then (0.9881)
8. service.Not (0.9880)
9. service.The (0.9880)
10. service.Just (0.9879)

Top 10 Opposite Words to "service":

1. ? (0.2194)
2. \U00100065 (0.2044)
3. Αθήνα (0.2015)
4. \U00100042 (0.2006)
5. ZAll (0.1975)
6. May-01 (0.1973)
7. \U00100071 (0.1955)
8. 2014X (0.1955)
9. Χωρίς (0.1950)
10. \U00100068 (0.1942)

Top 10 Similar Words to "work":

1. work-work (0.9840)
2. workwork (0.9828)

3. work.Could (0.9568)
4. work.That (0.9559)
5. work.Where (0.9548)
6. work.But (0.9548)
7. work.Since (0.9542)
8. work.And (0.9540)
9. work.Well (0.9536)
10. work.So (0.9536)

Top 10 Opposite Words to "work":

1. Tī (0.2713)
2. Fī (0.2631)
3. ĩ (0.2623)
4. ěŽě (0.2613)
5. Rī (0.2611)
6. 호스팅 (0.2600)
7. фрукты (0.2552)
8. .è (0.2534)
9. 30ä (0.2518)
10. Vī (0.2504)

## Conclusion & Comments

Findings for "Service":

- The **custom model** captures variant forms of "service" with decent semantic closeness, but many results appear as spelling variants or artificial constructions.
- The **pretrained model** shows common misspellings and variations of "service", but includes formatting-based tokens like "service.The".
- The **updated pretrained model** displays highly similar phrases with very high confidence scores, suggesting possible overfitting to Yelp-specific contexts. Opposite word outputs remain weak and include non-semantic or symbolic tokens across all models.

Findings for "Work":

- The **custom model** performs relatively well, returning synonyms related to employment and professionalism. Opposites, however, are largely unrelated or nonsensical.
- The **pretrained model** has better synonym recognition but struggles with irrelevant opposite terms.
- The **updated pretrained model** shows exceptional precision in similar word predictions but again suffers in producing meaningful opposite terms.

Model Comparison Summary:

Model	Similar Word Quality	Opposite Word Quality	General Comments
Custom Yelp Model	Moderate (varied forms)	Poor (random associations)	Needs more data, opposites unreliable
Pretrained FastText	Decent (misspellings, mix)	Weak (irrelevant tokens)	Good for synonyms, lacks antonym detection
Updated Pretrained	Decent (misspellings, mix)	Unstable (symbols, artifacts)	Best for similarity, opposites need major filtering