



What the Project Provides

1. Marathi Language Dataset for-
 - a. NER
 - b. Semantic Analysis
 - c. News Categorization
2. Open-Source Language Library for Marathi for Python
3. Named Entity Recognition for Marathi data
4. Sentiment Analysis for Marathi data
5. News Text Classification for Marathi data
6. Android Application with embedded NER, SA, NC, and OCR capabilities



Data Scraped From:



लोकमत



News Text Classification

अमर्याद

बातमी वर्गीकरण

अमर्याद

बातमी वर्गीकरण

वर्ग शोधा

लुका चुप्पी कमी झाले
कलेक्शन

वर्ग शोधा

बातमी मथळा: लुका चुप्पी कमी झाले
कलेक्शन
वर्ग: मनोरंजन



	Baseline- ULMFit	Our model using BERT
Test Accuracy	0.8175	0.9341

Number of training sentences: 9,673

	headline	label
0	सचिन तेंडुलकरने बाळासाहेबांच्या आठवणींना दिला ...	state
1	बहुजन वंचित आघाडीची पहिली यादी जाहीर, प्रकाश आ...	state
2	गृहमंत्रिपद हे पार्ट टाइम नाही, राजीनामा द्या,...	state
3	पुण्यात आघाडीचा पोपट पुन्हा जिवंत !	state
4	शरद पवार यांनी मिलिंद नावेंकर यांच्या कानात क...	state

Translation:

News Text: Lukka Chuppi's
Collection Decreases

Category: Entertainment

Sentiment Analysis

अमर्याद

भावना विश्लेषण

वर्ग शोध

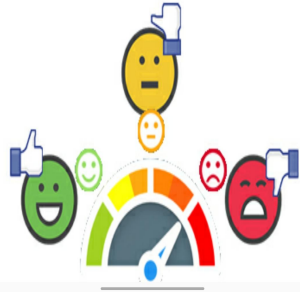
अमर्याद

भावना विश्लेषण

बागी सिनेमाच्या श्रृंखलेतील यादांचा 'बागी ३' हा सिनेमा कथारूपी अत्यंत कमजोर आहे.

भावना शोध

वाक्य: बागी सिनेमाच्या श्रृंखलेतील यादांचा 'बागी ३' हा सिनेमा कथारूपी अत्यंत कमजोर आहे.
भावना: नकारात्मक



	Baseline-Word2Vec	Our model using BERT
Test Accuracy	0.4897	0.6442

	index	Split	sentTag
56625	3348	सर्वे प्रवाशांना तसेच लष्कराच्या जवानांना इच्छित स्थळी सुरक्षित पोहोचवण्या	1
56626	3348	तसेच अडकलेल्या प्रवाशांसाठी हवाई दलातर्फे व्यवस्था केल्याची माहितीही	0
56627	3349	संत नाशकात १९६०च्या दशकात बाबांचा सत्कार झाला, त्यानिमित्ताने कवी कुं	1
56628	3349	खुद्द कुसुमाग्रजांनी ती कावेता वाचली, तेव्हा दिवे गेले	0
56629	3349	काही क्षणातच पेट्रोमॅक्सच्या बत्त्या व्यासपीठावर आल्या आणि त्या उजेडात ए	0
56630	3349	पुढे स्वतः बाबा बोलायला उभे राहिले आणि दिवे आले	0
56631	3349	साधुसंत येती घरा तोचि दिवाळी दसराआज आहे महोत्सव माझ्या शब्दांचा उ	-1
56632	3349	आकाशाचे मन कळत नाही वारा होऊन मुक्त झाल्याशिवाय प्रकाशाच्या देव	-1
56633	3349	लोहाराच्या भट्टीतील एखादी कलंदर ठिणगी झेप घेते आकाशात हरवलेपणाने	0
56634	3349	तसेच आहे माझ्या शब्दांचे क्षणाला माखलेले हे मोठेपण	1

Translation:

News Text: Baagi 3 of Baagi movie series has an extremely weak storyline.

Sentiment: Negative

Named Entity Recognition

अमर्याद

नामांकित संस्था ओळख

अमर्याद

नामांकित संस्था ओळख

बिल गेट्स यांनी दिला
मायक्रोसॉफ्टच्या संचालक
मंडळाचा राजीनामा.

वर्ग गोधा

वर्ग गोधा

वाक्य: बिल गेट्स यांनी दिला
मायक्रोसॉफ्टच्या संचालक मंडळाचा
राजीनामा.
शब्द -> वर्ग :
बिल -> व्यक्ती ; गेट्स -> व्यक्ती ;
मायक्रोसॉफ्टच्या -> संस्था



4322	142मिल राजकीय प्रेरणा पौटर अभिनय प्रेरणा राजेश अमिताभ बच्चन प्रेरणा राजा राजा अमर र	[ORG, ORG, PER, PER, PER, PER, PE
4323	142रिचर्ड बर्टन ओ टूलने चित्रपटापासून ऋषिकेश मुखर्जी अभिनित 'नमक हुराम' बनवला	[PER, PER, PER, PER, O, PER, PER,
4324	142थोमवर टीएल इलियट 'मर्डर कॅथड्रल' पथनाट्य लिहिले	[O, PER, PER, O, O, O, O]
4325	142ज्ञात राजाकडे चर्चकडे राजकुमार हेन्री सिंहासनावर विराजमान बालमित्राला चर्चचा नेमले	[O, O, LOC, PER, PER, O, O, O, LO
4326	142इरादा मिळवण्याचा चर्चच्या बसताच बालमित्राला देवाची चाकरी चर्चचे मित्रासोबत युद्धही	[O, O, LOC, O, O, PER, O, LOC, O,
4327	142एस इलियट रचनेतील टॅपटर दृश्याची स्तुती बॅकेटकडे टॅपटर पाठवतो	[PER, PER, O, O, O, O, PER, O, O]
4328	142मागितलेली ऐशोआरामाची साथने अमाप जमिनीचा बॅकेट नाकारतो	[O, O, O, O, O, O, PER, O]
4329	142चोथा टॅपटर म्हणतो बॅकेटला राजाचे ऐकल्यास हत्याच शहीद जन्मदिनी निर्वाणदिनी प्रलोभन	[O, O, O, PER, O, O, O, O, O, O, O
4330	142ऋषिकेश मुखर्जी अमिताभला मिलच्या मालकांची मित्राला म्हणतो, बनावे	[PER, PER, PER, O, O, O, O, O]
4331	142बनवतात	[O]
4332	142वेतनवाढ बोनससाठी मालकाशी लढतात	[O, O, O, O]
4333	142बॅकेटमध्ये ओ टूलच्या अभिनयापेक्षा रिचर्ड बर्टनच्या अभिनयाचे नमक हुराममध्ये खन्नाच्या अ	[PER, PER, PER, O, PER, PER, O, O]

Translation:

News Text: Bill Gates has resigned from board of Directors for Microsoft

Word -> Category

Bill -> PER; Gates -> PER; Microsoft -> ORG;