

Тема проекта:
Идентификация реплик персонажа
японской визуальной новеллы
с использованием методов машинного обучения
(на примере Юмэми из *Planetarian*)

Рыбакова Екатерина
2025 г.

Раздел 1: Вводная информация

Задача проекта и материалы

- ★ Построить модель бинарной классификации, которая по тексту реплики будет предсказывать, принадлежит ли она персонажу Юмэми.
- ★ **Материалы:** файл с размеченными по ролям репликами первой визуальной новеллы (обучающая/тестовая выборки); файл с текстом второй визуальной новеллы (валидационная/тестовая выборки).
- ★ Перенос модели на новые данные (transfer shifting).



Planetarian: Chiisana Hoshi no Yume (2004),
Planetarian: Snow Globe (2021).



「わたしは、廉価版の5000シリーズなので、頭部のインフォメーションリボンの特注オプションとなっています」

28

▲ ページ巻き戻し

▼ ページ順送り

文字領域の消去

しおりをはさむ

目次/しおりから

オートモード

1260 Yumemi, dialogue, 「争いの時代はおわりました。世界じゅうの人々が、手を取りあって、宇宙を目指す時がきたのです」
1261 Protagonist, narration, はるかな過去から現在へ、彼女の声はかろやかに時をわたって響く。
1262 Yumemi, dialogue, 「高みへ、はるかな高みへ…」
1263 Protagonist, narration, 超国家衛星通信網、宇宙往還機システム、国際宇宙港建設、月面再来訪計画。
1264 Yumemi, dialogue, 「光の空を越えて、夜の闇を抜けて…」
1265 Protagonist, narration, 挑戦の過程でおこった、避けられない事故の数々。
1266 Yumemi, dialogue, 「たくさんの涙の粒を、あたらしい夢のしずくにかえて…」
1267 Protagonist, narration, そして、人類の英知を結集した有人火星探査計画。
1268 Yumemi, dialogue, 「ついに人は、はるかに遠い惑星にさえ降り立ったのです」
1269 Protagonist, narration, 光でできた指示標が、南天に赤く輝く星を指す。俺の脳裏で、赤茶けた火星の地表が衍
1270 Yumemi, dialogue, 「これが、火星です」
1271 Yumemi, dialogue, 「直径 6 7 9 4 キロメートル、公転周期 6 8 7 日、平均気温マイナス 5 5 度、平均気圧 7 ヘク
1272 Yumemi, dialogue, 「この赤い星の表面に運河らしき痕跡が発見されると、人類はそこから、隣人の姿や物語を想い描
1273 Yumemi, dialogue, 「それは、高度に発達した科学を持つ異形の侵略者であり、美しいお姫様との英雄譚であり、調和
1274 Yumemi, dialogue, 「今では火星は、かつて生命が存在した痕跡だけを持つ、死の星であることがわかっています」
1275 Yumemi, dialogue, 「ですが、火星の運河を元に綴られたたくさんの物語は、今も色褪せることなく、人々を惹きつけてや
1276 Protagonist, narration, 彼女はそこで、しばし言葉を待った。「人類初の有人火星探査が成功したその年に、当
1277 Yumemi, dialogue, 「それ以来、本当にたくさんの方々が、ここで投影をご覧になりました」
1278 Yumemi, dialogue, 「この火星の姿は、まさに宇宙開発の成果とその象徴です」
1279 Yumemi, dialogue, 「そして同時に、戦いの神、マルスの象徴でもあります」

1 — txt-файл с размеченным по ролям текстом
Planetarian: Chiisana Hoshi no Yume,
740 “Yumemi” vs 2020 “not Yumemi”
(после препроцессинга)

733	Yumemi	dialogue	『失礼ですが、ススムさまのお歳はおいくつですか？』
734	Other	dialogue	『9歳』
735	Yumemi	dialogue	『そうしますと、残念ながら結婚は許可されないかと考
736	Narrator	narration	しばらくの沈黙。
737	Other	dialogue	『ゆめみちゃん、10年待ってよ！僕、10年したら迎えにき
738	Other	dialogue	『僕、それまでガマンするから。そうしたらケッコンし
739	Yumemi	dialogue	『それは、婚約ということでしょうか？』
740	Other	dialogue	『そうだよ、コンヤク！』
741	Yumemi	dialogue	『ですが、わたしには業務がありますので』
742	Other	dialogue	『だれも見えてない時に外に出ればいいんだよ、仕事のフ
743	Other	dialogue	『そうすれば僕、迎えに行くから。ゼッタイ、ゼッタイ
744	Narrator	narration	ためらうような数秒の後、ゆめみの声ははっきりと言っ
745	Yumemi	dialogue	『承知しました。重要命令として登録します』
746	Other	dialogue	『うんっ。ヒミツだよ、ヒミツの約束だよ...』
747	Narrator	narration	そこで映像は途切れ、また最初のシーンに戻る。
748	Narrator	narration	ちいさな男の子と女の子が交わした、無邪気で純粋な約
749	Narrator	narration	それは雪圏球に封じ込められた雪の欠片ばように、いつ
750	Narrator	narration	たとえ男の子は、全部忘れてしまったとしても。
751	Narrator	narration	里美は考えてしまう。
752	Satomi	thoughts	(忘れてしまう人間と、忘れないゆめみ...)

2 — xlsx-таблица с размеченным по ролям текстом
Planetarian: Snow Globe,
68 “Yumemi” vs 719 “not Yumemi”
(после препроцессинга)

Чем речь Юмэми уникальна?



- ★ Юмэми Хосино — сервисный робот-гид, работающий в планетарии. Её речь отличается вежливостью (*кэйго*), искренней доброжелательностью, и оптимизмом, что создаёт яркий контраст с усталым, циничным тоном главного героя первой игры.
- ★ Реплики Юмэми часто повторяются.

Этапы проекта: работа с первым текстом

- 1) Предобработка текста первой игры.
- 2) Векторизация текста: **MiniLM, Sonoisa, TohokuBERT**; выбор лучшей эмбеddинг-модели для задачи классификации.
- 3) Обучаем базовую модель **логистической регрессии** для бинарной классификации по эмбеddингам (первая игра: обучающая выборка 80%, тестовая выборка 20%).
- 4) Помимо логистической регрессии, в отдельном файле показана работа более сложных моделей: **случайного леса** и **градиентного бустинга (LightGBM)**.

Этапы проекта: работа со вторым текстом

- 5) Предобработка текста второй игры.
- 6) Векторизация текста: **TohokuBERT**.
- 7) Применяем обученную на тексте первой игры базовую модель **логистической регрессии**. Смотрим, как хорошо она отличает “Yumemi” от “not Yumemi”. 10% — валидационная выборка для определения оптимального порога вероятности. Оставшиеся 90% — тестовая выборка со скрытыми метками.
- 8) Помимо логистической регрессии, в отдельном файле показана работа более сложных моделей: **случайного леса** и **градиентного бустинга (LightGBM)**.
- 9) Выводы.
- 10) Краткий анализ файла с ошибками.

Дисбаланс классов

- ★ Коррекцию дисбаланса классов мы осуществляли с помощью встроенных параметров моделей: логистической регрессии (`class_weight`), случайного леса (`class_weight`) и градиентного бустинга LightGBM (`is_unbalance=True`), без использования методов оверсэмплинга или андерсэмплинга.
- ★ Судя по результатам тестов, умеренный дисбаланс классов в ($\approx 2.73:1$) не помешал работе моделей благодаря применению встроенных параметров и использованию адекватных эмбедингов.

Раздел 2: Реализация проекта (первая визуальная новелла)

1. Предобработка текста первой игры

- ★ На этапе предобработки были оставлены только символы японских слоговых азбук и иероглифы, пустые строки удалены.
- ★ Поскольку традиционные методы обработки текста (токенизация, n-граммы, синтаксический разбор) показали низкую эффективность на наших данных, мы выбрали подход на основе **эмбеддингов**, который позволяет моделям захватывать контекст и семантику без опоры на предварительное разбиение текста.

```
# Оставляем только японские символы, удаляем пунктуацию, цифры и т.д.  
def clean_text(text):  
    text = re.sub(r'^\u3040-\u309F\u30A0-\u30FF\u4E00-\u9FFF]', '', text)  
    return text
```


Почему традиционные методы плохо работают с японским текстом:

- ★ В японском письме нет явных границ между словами → токенизаторы часто ошибаются при разбиении текста.
- ★ Грамматическая информация выражается через набор суффиксов, частиц и вспомогательных глаголов, часто «приклеенных» к основам → n-граммы не понимают их значений и легко теряют смысл.
- ★ Вежливость в японском — это устойчивые глагольные конструкции, которые могут состоять из нескольких морфем → простая токенизация разрывает связные выражения, и модель не улавливает контекст.
- ★ N-граммы не различают стили речи и теряют обобщаемость.

2. Векторизация текста (эмбеддинги)

Мы обратились к предобученным эмбеддинг-моделям, которые используют собственные токенизаторы. Мы опробовали три модели:

- ★ paraphrase-multilingual-MiniLM-L12-v2 (из sentence-transformers);
- ★ sonoisa/sentence-bert-base-ja-mean-tokens (из sentence-transformers);
- ★ cl-tohoku/bert-base-japanese (PyTorch).

При обучении модели логической регрессии на первом тексте третья эмбеддинг-модель показала наилучший результат.

2. Сравнение эмбединг-моделей

==== MiniLM ====

	precision	recall	f1-score	support
not Yumemi	0.92	0.88	0.90	404
Yumemi	0.71	0.80	0.75	148
accuracy			0.86	552
macro avg	0.82	0.84	0.83	552
weighted avg	0.87	0.86	0.86	552

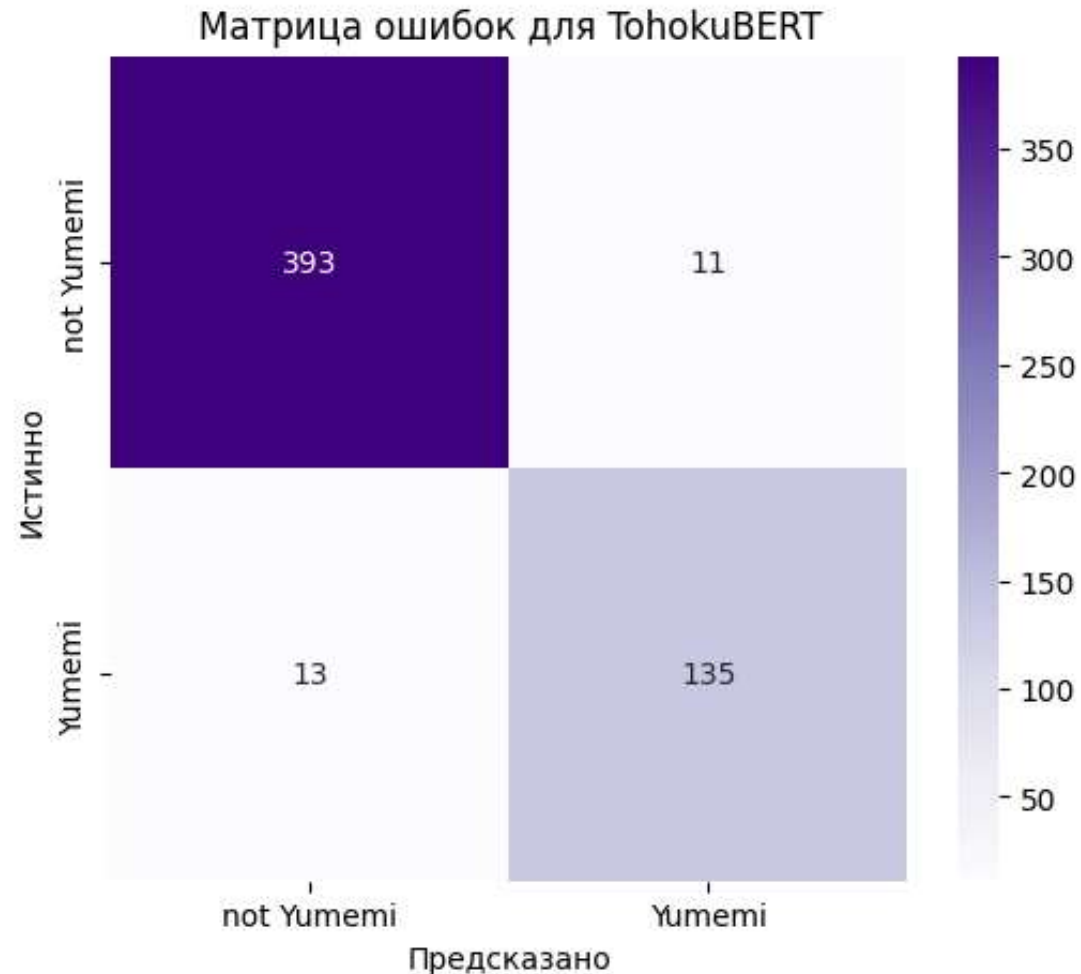
==== Sonoisa ====

	precision	recall	f1-score	support
not Yumemi	0.94	0.91	0.92	404
Yumemi	0.77	0.84	0.80	148
accuracy			0.89	552
macro avg	0.85	0.88	0.86	552
weighted avg	0.89	0.89	0.89	552

==== TohokuBERT ====

	precision	recall	f1-score	support
not Yumemi	0.97	0.97	0.97	404
Yumemi	0.92	0.91	0.92	148
accuracy			0.96	552
macro avg	0.95	0.94	0.94	552
weighted avg	0.96	0.96	0.96	552

2. Наилучший результат у TohokuBERT



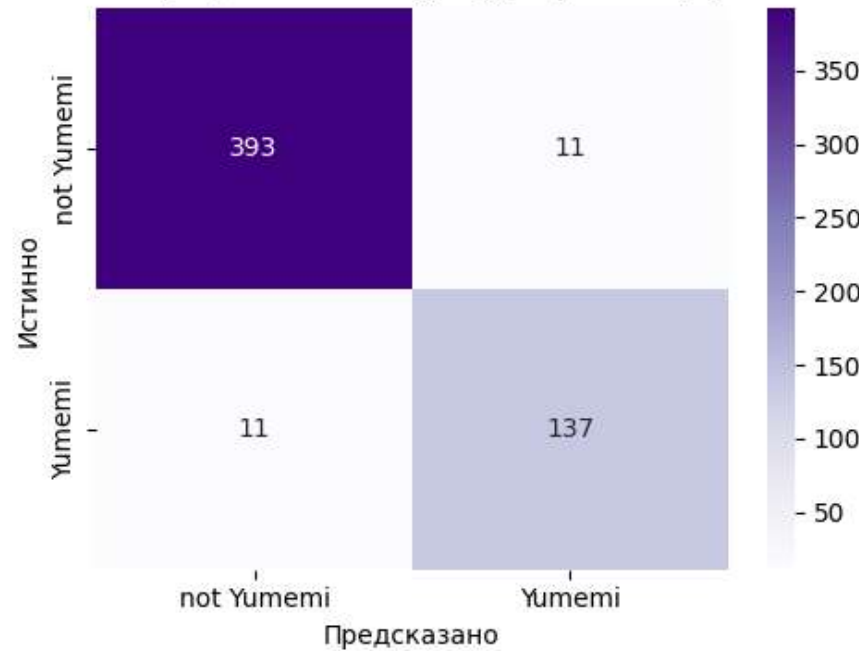
- ★ F-мера для “Yumemi” — 0.92.
- ★ Эта модель наиболее эффективно улавливает контекст, грамматику и стилистические нюансы языка.
- ★ Высокие показатели точности и полноты делают её оптимальным выбором для решения нашей задачи.

3. Алгоритм работы классификационных моделей

- 1) Подготавливаем данные: загружаем датасет с эмбедингами и исходными метками, создаём бинарную целевую метку (label), извлекаем признаки X (эмбединги) и целевую переменную y .
- 2) Делим данные на train/test (80/20) с параметром stratify= y .
- 3) Обучаем модель с кросс-валидацией для оценки F-меры, (опционально) подбираем оптимальный порог вероятности на train (перебор от 0 до 1, $by=0.01$), чтобы максимизировать F-меру. Задаём параметры с учётом дисбаланса.
- 4) Сохраняем обученную модель в файл (pickle).
- 5) Оцениваем модель на тестовой выборке, выводим классификационный отчёт и матрицу ошибок. Лучшую модель выбираем по метрикам миноритарного класса.

3. Логрег + TohokuBERT (оптимальный порог = 0.47)

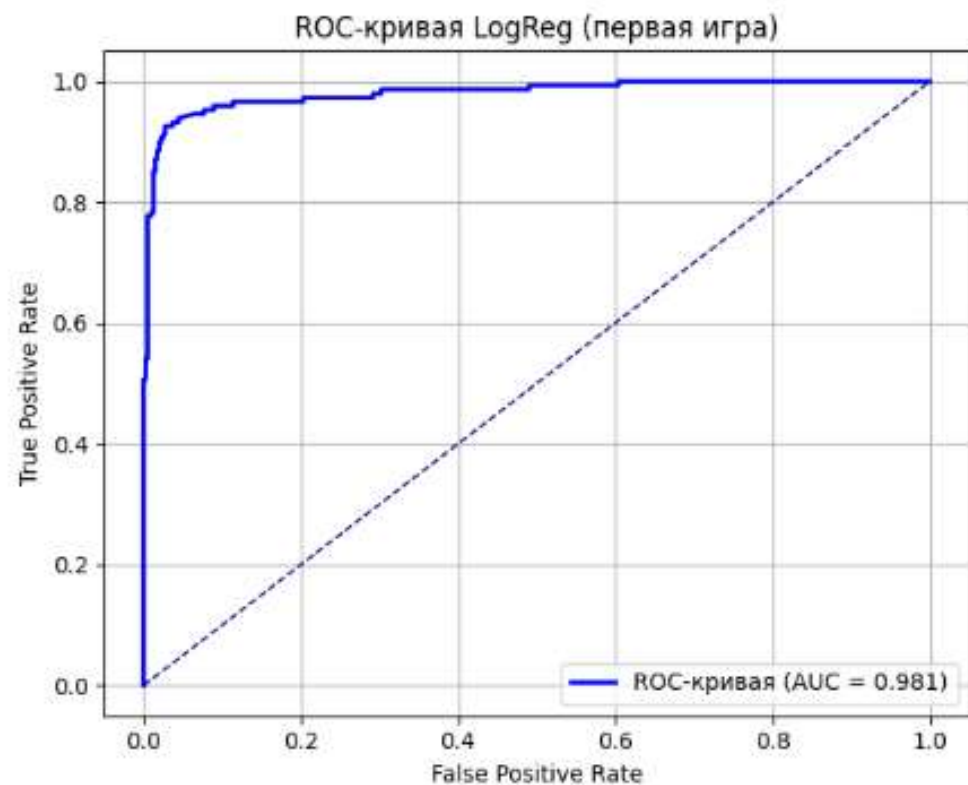
Матрица ошибок LogReg (первая игра)



Классификационный отчёт LogReg (первая игра):

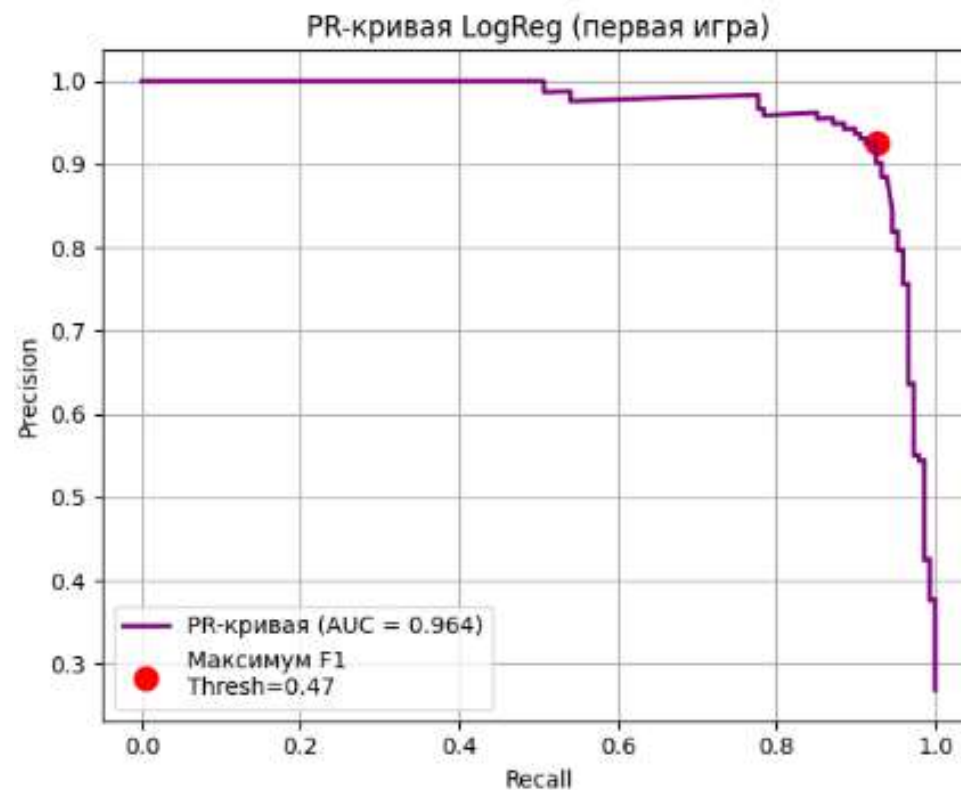
	precision	recall	f1-score	support
not Yumemi	0.97	0.97	0.97	404
Yumemi	0.93	0.93	0.93	148
accuracy			0.96	552
macro avg	0.95	0.95	0.95	552
weighted avg	0.96	0.96	0.96	552

- ★ Модель показывает высокую точность: общая accuracy — 96%.
- ★ Класс “Yumemi” определяется хорошо: F-мера = 0.93.
- ★ Ошибки: 11 FN и 11 FP.
- ★ Хороший баланс между precision и recall для обоих классов.

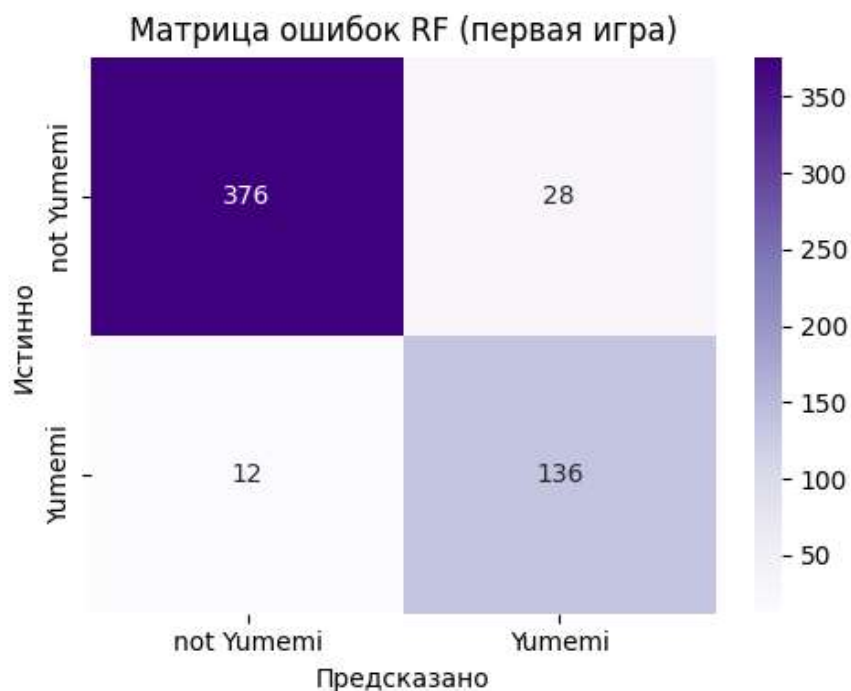


ROC AUC = 0.981 — модель отлично отличает положительный класс от отрицательного.

PR AUC = 0.964 — модель хорошо работает с положительным классом, высокая точность при высокой полноте.



4. Случайный лес + TohokuBERT (опт. порог = 0.26)

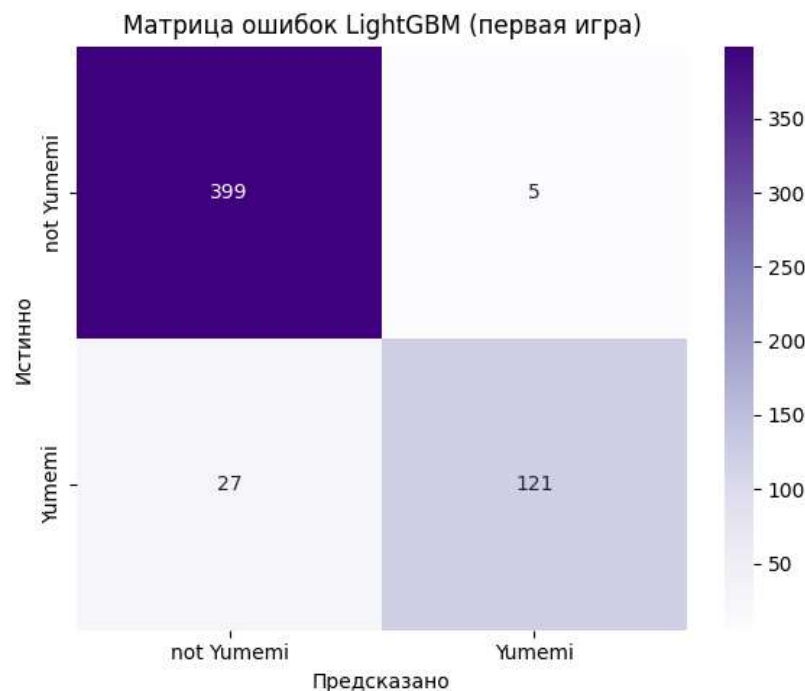


Классификационный отчёт RF (первая игра):

	precision	recall	f1-score	support
not Yumemi	0.97	0.93	0.95	404
Yumemi	0.83	0.92	0.87	148
accuracy			0.93	552
macro avg	0.90	0.92	0.91	552
weighted avg	0.93	0.93	0.93	552

- ★ Без подбора гиперпараметров. GridSearchCV и RandomizedSearchCV показали результаты хуже, как и отсутствие оптимального порога.
- ★ F-мера на кросс-валидации (5 фолдов): 0.768 ± 0.012
- ★ Логрег показал результаты лучше по всем основным метрикам.

4. LightGBM + TohokuBERT (стандартный порог = 0.5)



Классификационный отчёт LightGBM (первая игра):

	precision	recall	f1-score	support
not Yumemi	0.94	0.99	0.96	404
Yumemi	0.96	0.82	0.88	148
accuracy			0.94	552
macro avg	0.95	0.90	0.92	552
weighted avg	0.94	0.94	0.94	552

- ★ F-мера на кросс-валидации (5 фолдов): 0.876 ± 0.018 .
- ★ LightGBM показывает высокую точность на “Yumemi”, но recall ниже, чем у логрег — модель осторожнее, но пропускает больше “Yumemi”.
- ★ Логрег всё равно показывает более сбалансированный результат.

Раздел 3: Реализация проекта (вторая визуальная новелла)

5. Предобработка текста второй игры;

6. Генерация эмбедингов TohokuBERT

```
character
Narrator      420
Satomi        195
Other         83
Yumemi        68
Goro          21
Name: count, dtype: int64
```

Генерация эмбедингов: 100%|██████████| 78/78 [00:04<00:00, 18.80it/s]

Генерация эмбедингов: 100%|██████████| 709/709 [00:34<00:00, 20.53it/s]

- ★ Для того, чтобы улучшить результаты работы модели логической регрессии, обученной на репликах из первой игры, мы нашли оптимальный порог вероятности по F-мере на валидационной выборке: 0.85 ($F1 = 0.824$). 10% — валидационная выборка, остальные 90% — тестовая без разметки.
- ★ Разделение потребовалось, чтобы избежать утечки данных.

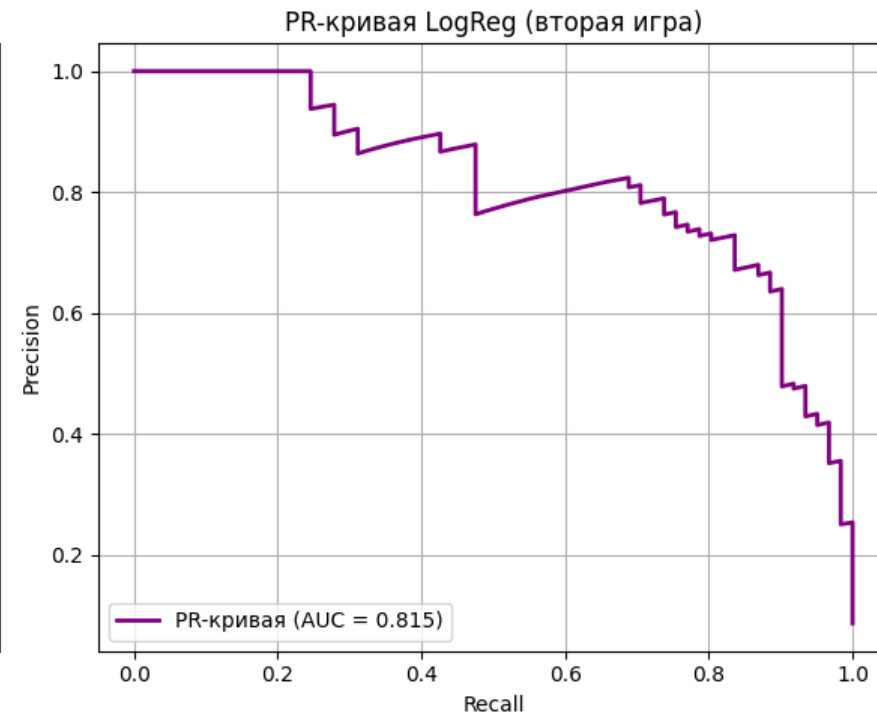
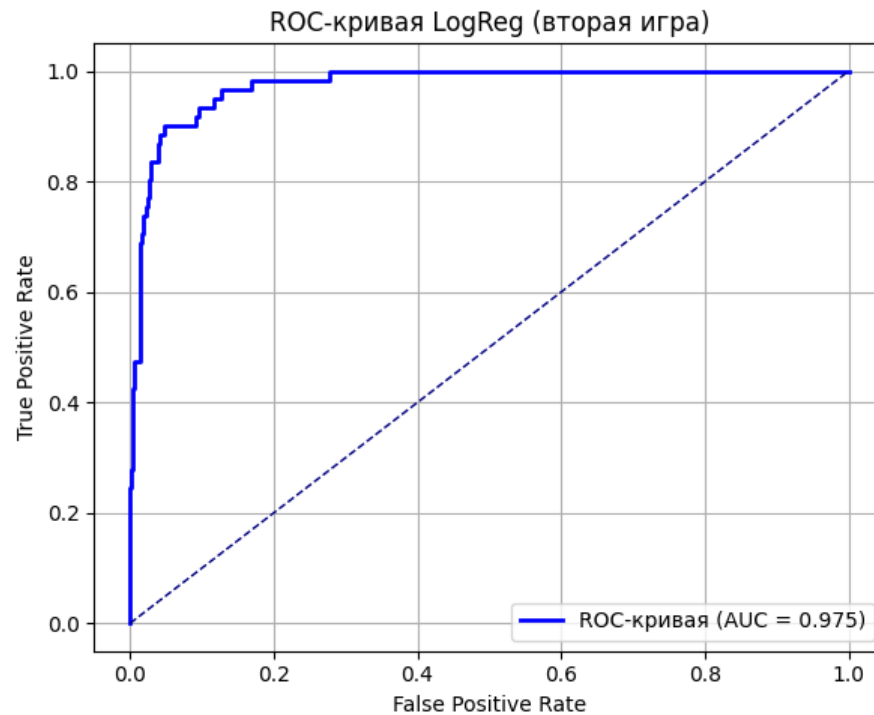
7. Логистическая регрессия (опт. порог 0.85)

Классификационный отчёт LogReg (вторая игра):

	precision	recall	f1-score	support
not Yumemi	0.99	0.96	0.97	648
Yumemi	0.67	0.89	0.76	61
accuracy			0.95	709
macro avg	0.83	0.92	0.87	709
weighted avg	0.96	0.95	0.96	709

- ★ Мало FN, что важно для задачи идентификации реплик.
- ★ Высокий recall в ущерб precision.
- ★ Общая точность 0.95.





- 1) ROC-кривая слабо зависит от дисбаланса классов (FPR считает только негативные классы).
- 2) PR-кривая очень чувствительна к дисбалансу классов. Кривая выглядит «рваной» из-за того, что precision падает при росте recall (положительного класса мало).

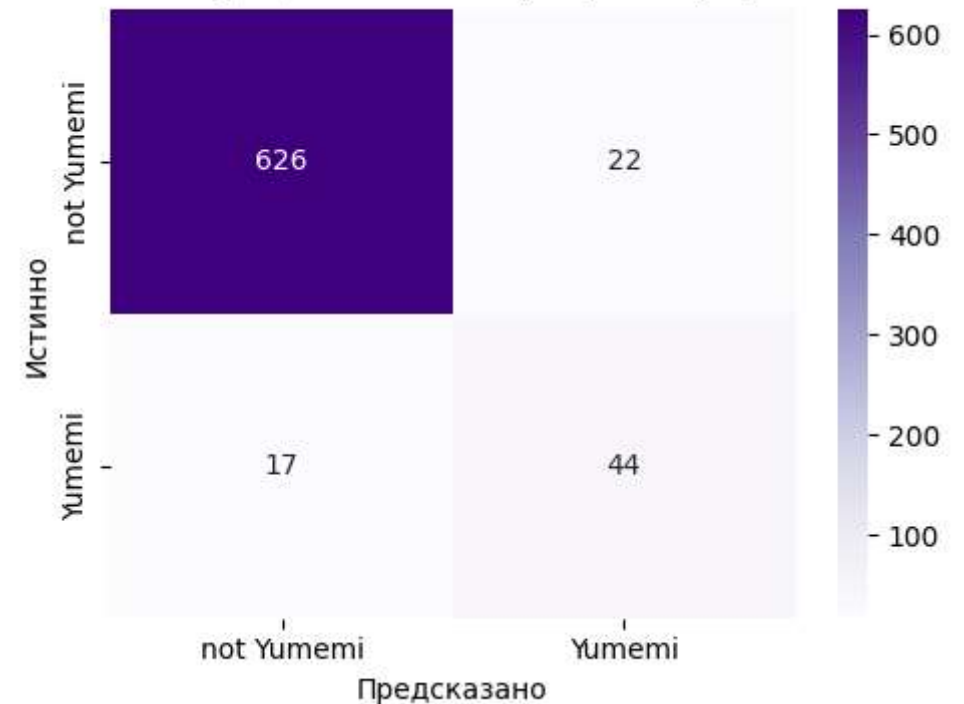
7. Случайный лес (опт. порог 0.46)

Классификационный отчёт RF (вторая игра):

	precision	recall	f1-score	support
not Yumemi	0.97	0.97	0.97	648
Yumemi	0.67	0.72	0.69	61
accuracy			0.94	709
macro avg	0.82	0.84	0.83	709
weighted avg	0.95	0.94	0.95	709

- ★ Случайный лес меньше ошибается с FP, но пропускает больше положительных — хуже recall и F-мера для “Yumemi”.
- ★ Логрег показывает лучший баланс между recall и ассурасу (F-мера) для класса “Yumemi”, лучше улавливает положительные случаи (меньше FN), но чуть больше ошибочных срабатываний (FP).

Матрица ошибок RF (вторая игра)



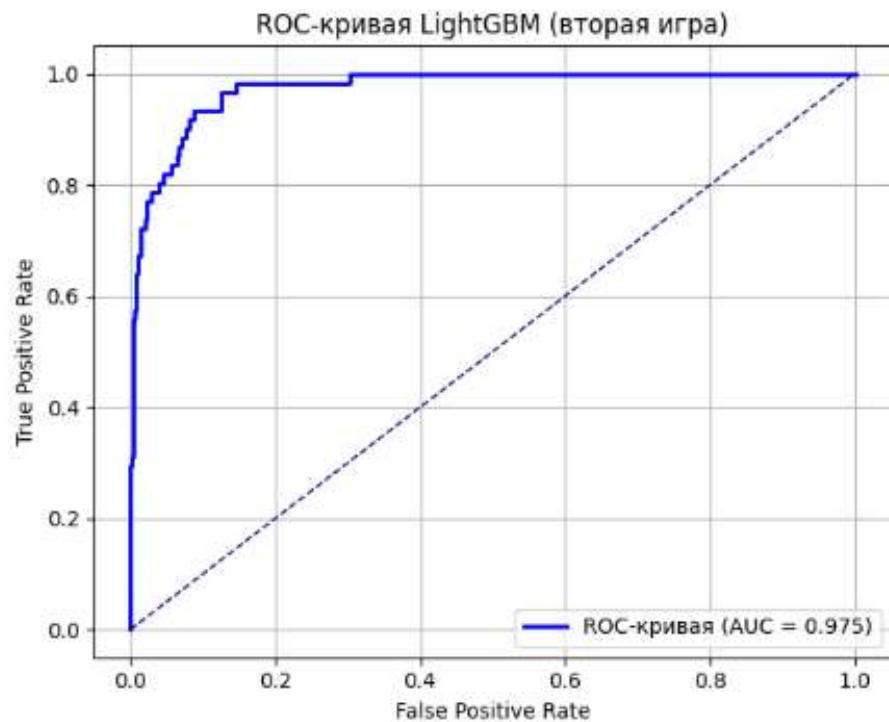
8. Градиентный бустинг LightGBM (опт. порог 0.96)

Классификационный отчёт LightGBM (вторая игра):

	precision	recall	f1-score	support
not Yumemi	0.98	0.98	0.98	648
Yumemi	0.77	0.75	0.76	61
accuracy			0.96	709
macro avg	0.87	0.87	0.87	709
weighted avg	0.96	0.96	0.96	709

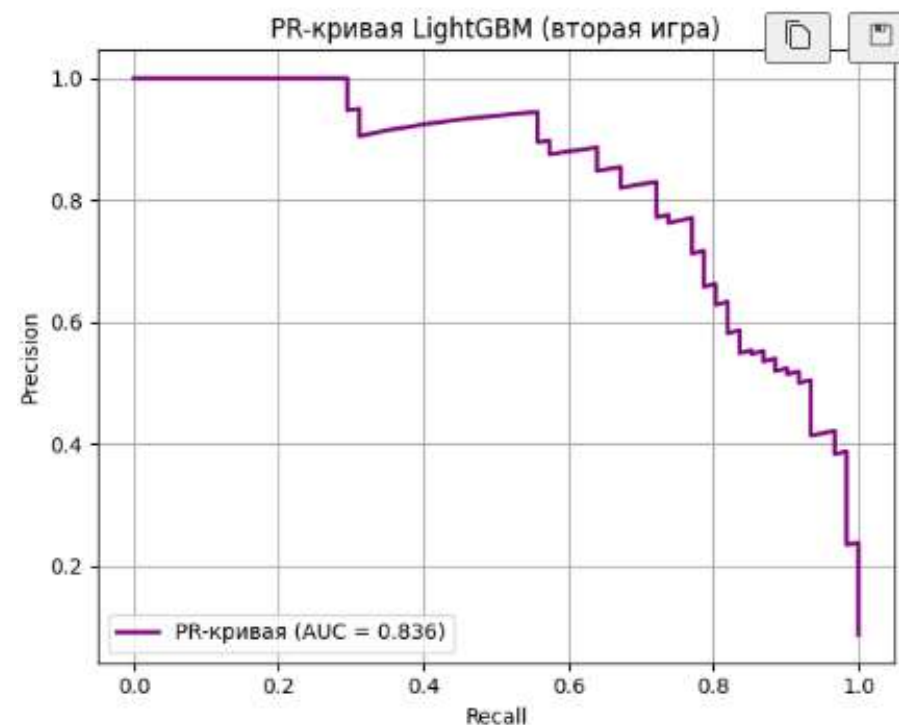
- ★ **LightGBM.** Accuracy: 0.96, F1 (Yumemi): 0.76, Recall (Yumemi): 0.75, FP: 14, FN: 15
- ★ **Логрег.** Accuracy: 0.95, F1 (Yumemi): 0.76, Recall (Yumemi): 0.89, FP: 27, FN: 7
- ★ LightGBM даёт большую точность и меньше FP, однако пропускает больше FN.





ROC AUC = 0.975: модель хорошо различает классы по вероятностным показателям.

PR AUC = 0.836 (у логрег был 0.815),
прирост в 2.6%. Качество ранжирования
немного улучшилось.



9. Выводы

- ★ Все модели отлично определяют мажоритарный класс как в первой, так и во второй игре.
- ★ LightGBM показал сбалансированные метрики: precision, recall и F-меру выше 0.75, что свидетельствует о хорошей способности модели как выявлять реплики Юмэми, так и не пропускать их значительную часть. Допущено меньше всего ошибок ($FP + FN = 29$).
- ★ На наш взгляд, логрег и LightGBM показали достойные результаты.

Модель	Precision (Yumemi)	Recall (Yumemi)	F1-score (Yumemi)	False Positives	False Negatives	Итог
Logistic Regression	0.67	0.89	0.76	27	7	Высокий recall, много FP
Random Forest	0.67	0.72	0.69	22	17	Средние показатели
LightGBM	0.77	0.75	0.76	14	15	Лучший баланс, меньше FP

9. Выводы

- ★ Градиентный бустинг последовательно строит ансамбль слабых моделей, каждая из которых исправляет ошибки предыдущих. Это позволяет модели эффективно захватывать сложные зависимости и паттерны в данных, особенно при работе с высокоразмерными и неявными признаками, как в случае с BERT-эмбедами.
- ★ Модель логистической регрессии допустила меньше всего FN, что важно для задачи идентификации реплик, несмотря на большее общее количество ошибок (34), чем у градиентного бустинга (29).
- ★ Меньше FN → максимальное покрытие всех фраз Юмэми → минимальный риск пропустить реплику.
- ★ Случайный лес справился хуже, потому что переобучился на первую игру и не смог адаптироваться к изменениям стиля и контекста, в отличие от более устойчивых LightGBM и логистической регрессии.

10. Краткий анализ файла с ошибками LogReg

```
1 character,type,text,label,pred_label
2 Narrator,narration,その上,0,1 вежливость
3 Other,dialogue,主力商品が欲しいところですねできれば冬ならではの感じの,0,1 вежливость
4 Satomi,dialogue,あなたの今日の業務はなんだったかしら,0,1
5 Other,dialogue,痛い痛いもう降参降参参りました攻撃しないでっ,0,1 вежливость
6 Yumemi,dialogue,お元気そうで安心しました,1,0
7 Other,dialogue,僕それまでガマンするからそうしたらケツコンしよ,0,1
8 Other,dialogue,スノーグローブなんてどうでしょうガラスの中に雪が積もるアレです,0,1 вежливость
9 Other,dialogue,あのすみませんわたしぜんぜん対応できなくて,0,1 вежливость
10 Other,dialogue,倉橋さんすみません,0,1 вежливость
11 Goro,dialogue,現状ではハード絡みの可能性も捨てきれないです,0,1 вежливость
12 Other,dialogue,だって普通に考えたらわかりますよー最初だけみんなでチャホヤしといて飽きたら見向きもしないなんてわたしならぜったい辞めちゃいますもんねーゆめみちゃん,0,1 вежливость
13 Goro,dialogue,館長に報告どうします,0,1 вежливость
14 Yumemi,dialogue,ペーパーフィルター在庫は十分ですがコーヒー豆の賞味期限まで一週間を切りましたので,1,0 вежливость
15 Other,dialogue,年分をフルチェックですか物理的に無理です,0,1 вежливость
16 Other,dialogue,ところでどうします原因は確定しましたし現在は既に発現しない不具合であることも自信を持って言えますが,0,1 вежливость
17 Satomi,dialogue,景気が悪いことは考えないようにしましょうお客が逃げちゃう,0,1 вежливость
18 Yumemi,dialogue,あのお聞きしてよろしいですか,1,0 вежливость
19 Other,dialogue,えいっえいっえいっ,0,1
20 Other,dialogue,そうすれば僕迎えに行くからゼツタイゼツタイ行くから,0,1
21 Satomi,dialogue,もうちょっと様子を見ましようせっかくの海外視察なんだから余計な心配させたくないでしよ,0,1
22 Yumemi,dialogue,結婚をした場合必ず退職しなければならないのでしょうか,1,0
23 Narrator,narration,花菱デパート本店屋上プラネタリウム館受け付けの券売機前,0,1
24 Satomi,dialogue,とりあえず恋人たちの前途に幸せ多かれと願っておくことにするわ,0,1
25 Satomi,thoughts,もしもいつかこの職場がなくなつたとしてもわたしがここで学んだこと感じたことの種を残すために,0,1
26 Yumemi,dialogue,お二人ともどうぞ末長くお幸せに,1,0
27 Satomi,dialogue,でも館長彼女は人ではないですよ,0,1 вежливость
28 Other,dialogue,秘密って言われたからってデバッグルーチン自力でブックして命令スタッフを不可視にするなんてよくまあ手の込んだことができたもんです,0,1 вежливость
29 Yumemi,dialogue,はいあの時はとても大変でした,1,0
30 Yumemi,dialogue,プラネタリウム館受け付け券売機前にて音声による顧客呼び込み及び業務案内です,1,0
31 Goro,dialogue,あれはものすごく限定された条件下でないと起こらないと立証されてますし今は例外行動チェックが厳しくなってますから,0,1 вежливость
32 Satomi,dialogue,ですが館長,0,1 вежливость
33 Satomi,dialogue,そうですそうです彼女が新人ではありません,0,1 вежливость
34 Other,narration,なかなかお会いできないので健康状態に問題がおりではないかと心配していました,0,1 вежливость
35 Goro,dialogue,了解です,0,1 вежливость
```


10. Краткий анализ файла с ошибками LightGBM

```
1 character, type, text, true_label, pred_label
2 Other, dialogue, 痛い痛いもう降参降参っ参りました攻撃しないでっ, 0, 1 вежливость (маиримасу)
3 Satomi, dialogue, ねえもうすぐ投影終わるからおねえさん仕事に戻りたいんだけど, 0, 1 вежливость? (онээсан)
4 Yumemi, dialogue, お元気そうで安心しました, 1, 0 неясно
5 Yumemi, dialogue, そうですかとても安心しました, 1, 0 неясно
6 Yumemi, dialogue, はい人間にはロボットにしか話せないことがあるから里美さんとふたりきりになった時に何も言わずに愚痴を聞いてあげてとおっしゃいました, 1, 0 неясно
7 Yumemi, dialogue, わかりましたそれで婚約は解消ということで承りました, 1, 0 неясно
8 Goro, dialogue, 館長に報告どうします, 0, 1 вежливость
9 Yumemi, dialogue, ペーパーフィルター在庫は十分ですがコーヒー豆の賞味期限まで一週間を切りましたので, 1, 0 неясно
10 Other, dialogue, 年分をフルチェックですか物理的に無理です, 0, 1 вежливость
11 Yumemi, dialogue, それは婚約ということでしょうか, 1, 0 неясно
12 Other, dialogue, ところでどうします原因は確定しましたし現在は既に発現しない不具合であることも自信を持って言えますが, 0, 1 вежливость
13 Yumemi, dialogue, あのお聞きしてよろしいですか, 1, 0 неясно
14 Satomi, dialogue, もうちょっと様子を見ましょうせっかくの海外視察なんだから余計な心配させたくないでしょ, 0, 1 неясно
15 Yumemi, dialogue, はいどうぞなんなりとお尋ねください, 1, 0 неясно
16 Yumemi, dialogue, 結婚をした場合必ず退職しなければならないのでしょうか, 1, 0 неясно
17 Satomi, dialogue, わかったわ通常業務に戻ってちょうだい, 0, 1 неясно
18 Yumemi, dialogue, はい話せば長くなるのですが, 1, 0 неясно
19 Yumemi, dialogue, わたしが表口前で呼び込み業務及び花菱デパート特産展チラシ配布業務に従事していましたところ, 1, 0 неясно
20 Satomi, dialogue, とりあえず恋人たちの前途に幸せ多かれと願っておくことにするわ, 0, 1 неясно
21 Satomi, thoughts, もしもいつかこの職場がなくなってもわたしがここで学んだこと感じたことの種を残すために, 0, 1 неясно (ватаси?)
22 Yumemi, dialogue, お二人ともどうぞ末長くお幸せに, 1, 0 неясно
23 Satomi, dialogue, わかるわよあとねおばさんじゃなくておねえさん, 0, 1 вежливость? (обасан, онээсан)
24 Other, dialogue, 秘密って言われたからってデバッグルーチン自力でブックして命令スタッフを不可視にするなんてよくまあ手の込んだことができたもんです, 0, 1 вежливость
25 Other, dialogue, すみませーんゆめみちゃん来ませんでしたかってなんかイヤなのがいる, 0, 1 вежливость
26 Yumemi, dialogue, はいあの時はとても大変でした, 1, 0 неясно
27 Yumemi, dialogue, プラネタリウム館受け付け券売機前にて音声による顧客呼び込む及び業務案内です, 1, 0 неясно
28 Satomi, thoughts, だいたいこういう無茶なお客さんのあしらひこそゆめみが一番得意にしてるところなのに, 0, 1 вежливость (окаякусан)
29 Yumemi, dialogue, 昨日みなさんはたまには里美さんの愚痴を聞いてあげてとおっしゃいました, 1, 0 неясно
30 Satomi, dialogue, そうですそうですが彼女は新人ではありません, 0, 1 вежливость
31
```

Спасибо за внимание!

