

Построение векторов личности по характеристикам человека с помощью языковых моделей.

Вступление.

Несколько лет назад в интернет-магазине книг я случайно увидел книгу Михаила Бородянского под названием "8 цветных психотипов: Кто вы? Векторная психология". Я приобрел данную книгу и прочитал ее. На основании прочитанного я сделал для себя определенные выводы о применимости векторной психологии в жизни. Размышляя, я смог по определенным признакам присвоить тот или иной вектор себе, своим знакомым и остановился на том, что действительно векторная психология "работает".

В своей книге М. Бородянский упоминает о том, что основной вклад в развитие векторной психологии внес петербургский психолог Виктор Константинович Толкачев.

Выбрав данную тему для своей работы, я ознакомился с книгой В. К. Толкачева "Роскошь системного самопознания".

В своей книге "Роскошь системного самопознания" В. К. Толкачев пишет такие строки: Векторная алгебра в психоанализе — это следующий этап теоретических рассуждений. Хватит ли жизни на исследование данной темы?

Краткое ознакомление с векторной психологией.

В своей работе В.К. Толкачев пишет о том, что всех людей можно разделить на восемь психологических типов, каждому типу соответствует чувствительное "отверстие" на теле, например, ухо, глаз, нос, рот и каждый тип испытывает тягу к своему цвету.

Психологический тип имеет условное название вектор и каждый вектор имеет соответствующий психологическому типу цвет (коричневый, черный, красный, оранжевый, желтый, зеленый, синий, фиолетовый).

Первый вектор, коричневый, разработал Зигмунд Фрейд более 100 лет назад.

Разделение людей по векторам основано на том, что в далекие времена для выживания первобытной стаи требовалось развитие различных качеств у ее представителей.

Например, синий вектор (ухо), обладает хорошим слухом, ложится спать в 5 утра (охраняет ночью), зеленый (глаз) обладает хорошим зрением, красный вожак, лидер, черный может выполнять длительное время монотонную физическую работу и т.д.

Ниже приведены некоторые отрывки из книги В. К. Толкачева "Роскошь системного самопознания": Каждый психологический тип испытывает бессознательную тягу к своему цвету. Цветовое воздействие на капсулу живого вещества — это механизм ее биохимической регуляции.

Внешняя среда не абсолютно определяет функциональный тип, но лишь мобилизует предрасположенности из имеющегося врожденного арсенала. Внешняя среда как бы только окультуривает тип, ментально модально окрашивая его.

Понятие типа есть некоторая идеальная научная категория, к точному описанию которой наука лишь приближается с той или иной степенью точности. И человеческий тип — это абстракция, идеальная научная категория, служащая для удобства описания многообразия объективных реальностей под названием «люди».

Типовой психосмысл, как отдельного индивида, так и целых народов и континентов, по сути, совпадают.

Вектор в неврозе — это характеристики реализованного вектора, но со знаком «минус». Он всегда сопровождается биохимическими проблемами, источник которых может быть как внутри организма, так и извне, будь то воспитание, культура и т. д.

Соответствующие черты характера, характерный комплекс, также умножается на -1. Отважный красный становится патологическим трусом. Аккуратный коричневый становится патологическим неряхой. Скардный оранжевый — транжирой и т. д. Зеленый (зрительный) слепнет. Желтый впадает в анорексию (полный отказ от пищи). Черный на очередные 33 года садится на печь. Синий глухнет. У фиолетового обратное состояние функции выжить, во что бы то ни стало, приобретает «большую» модальность — выжить лично мне, во что бы то ни стало!

Созидаемое — всегда психологический портрет созидателя!

Роскошь системного самопознания, навыки и умение практического использования системно-векторного психоанализа позволит нам:

- свободно выбирать собственные цели и приоритеты;
- точно определять и противостоять внешней манипуляции;
- создавать по настоящему близкие отношения;
- ощущать полноту жизни во всем ее многообразии;
- управлять своей жизнью и судьбой;
- реализовать свое функционально ролевое предназначение.

Основная сверхзадача данной теории — познай самого себя и исполни свое вселенское предназначение.

Применение языковых моделей на основе психологических векторов в современном мире.

Применение:

1. По прямому назначению, для чего и создавалась данная векторная система.

С помощью языковых моделей, представленных в данной работе, можно определять вектора человека по содержанию потребляемой информации, например, электронные книги, фильмы, музыка и т.п., а также по произнесенным словам при разговоре.

Определить вектора человека с целью развития врожденного потенциала, поиска путей выхода из жизненных кризисов и т.д.

Определить вектора ребенка и понять, как его воспитывать, чтобы он рос в гармонии с собой и окружающим миром. Например, “красного” ребенка нужно постоянно хвалить, по поводу и без повода, в то же время коричневого не стоит хвалить незаслуженно. Если у ребенка нет черного вектора, то нет смысла пытаться делать из него чемпиона по какому-либо виду спорта, не получится.

2. Построение рекомендательных систем.

Согласно утверждению из книги В.К. Толкачева “Созидаемое — всегда психологический портрет созидателя”, все, что сделано, нарисовано, написано и т. п. есть отображение внутреннего мира самого создателя, то есть созданное будет обладать теми же векторами, что и сам создатель. В своей книге В.К. Толкачев пишет о том, что люди с одинаковыми векторами понимают друг-друга и между людьми вовсе может не быть понимания, если их вектора различны, т.е. подобным векторам нравится подобное. Таким образом, вычислив, вектора определенного продукта и зная вектора некоторых людей, групп людей, можно порекомендовать продукт конкретному человеку, группе.

Можно решить проблему холодного старта рекомендательной системы, например, задавая при регистрации несколько ненавязчивых вопросов (любимый цвет, любимая геометрическая фигура, любимая цифра) и определив вектора нового пользователя.

3. Фильтрация потребляемого контента.

В режиме реального времени можно определять вектора и степень их принятия (позитивность) потребляемого контента и блокировать доступ к некоторым блокам данных при выявлении в них негативной информации.

4. Тестирование при приеме на работу.

Можно составить шаблон автобиографии при приеме на работу и автоматически определять основные вектора претендента на определенную должность, что позволит принять нужного человека и формировать максимально эффективную команду.

Из В.К. Толкачева:

Начальник фиолетового вектора — большая ошибка!

Сочетание двух векторов: фиолетового и оранжевого — это блестящие кабинетные ученые одиночки, которым свойственно логическо-алогическое, вполне академическое мышление.

Коричневый любит сочинять инструкции, черный любит выполнять инструкции, оранжевый любит их еще дополнять ГОСТами.

5. Наделение характером программных ботов.

Программный бот после непродолжительного общения может определить вектора собеседника и понять, что будет интересно человеку, а что нет, что может огорчить его или наоборот порадовать. Можно придать программе как бы характер, настроение, переменчивое поведение, но понятное человеку.

Построение языковых моделей.

Для построения языковых моделей из книг В. К. Толкачева и М. Бородянского выбрана информация и созданы текстовые файлы с полным описанием каждого вектора, а также файлы только с негативным описанием каждого вектора.

Загрузка описания психологических векторов.

Загружаем полное описание каждого вектора из подготовленных тестовых файлов.

```
person_vectors = ("brown", "black", "red", "orange", "yellow", "green", "blue", "purple")
```

```
df_vectors = pd.DataFrame([], columns=["description", "vectorId"])
```

```
# Загружаем описание каждого вектора из книги В.К. Толкачева
```

```
for person_vector in person_vectors:
    with open("vectors/" + person_vector + ".txt", encoding="utf8") as rf:
        for sentence in split_text(rf.read().strip()):
            if len(sentence) <= 2: continue
            df_vectors = df_vectors.append({"description": sentence, "vectorId": person_vectors.index(person_vector)})
```

```
# Загружаем описание каждого вектора из книги М.В. Бородянского
```

```
for person_vector in person_vectors:
    with open("vectors_/" + person_vector + ".txt", encoding="utf8") as rf:
        #sentences = re.sub(r'\s+', ' ', rf.read(), flags=re.M)
        #for sentence in re.split(r'(?<=[.!?]) ', sentences):
        for sentence in split_text(rf.read().strip()):
            if len(sentence) <= 2: continue
            df_vectors = df_vectors.append({"description": sentence, "vectorId": person_vectors.index(person_vector)})
```

Таким же образом загружаем негативное описание каждого вектора и убираем из полного описания негативное, получаем позитивное описание.

Таким образом у нас есть три набора данных:

- полное описание векторов (4562 строки), набор данных с именем df_vectors;
- чисто позитивное описание векторов (4184 строки), набор данных с именем df_vectors_pos;
- чисто негативное описание векторов (396 строки), набор данных с именем df_vectors_neg.

Полное описание используется для построения модели многоклассовой классификации с целью определения значений векторов, а позитивное и негативное описание для построения регрессионной модели (0. - негативное описание, 1. - позитивное) с целью определения степени позитивности векторов (8 регрессионных моделей).

Предобработка описания векторов.

```
# Токенизация
df_vectors["description"] = df_vectors["description"].apply(lambda txt: word_tokenizer.tokenize(txt))

# Удаление стоп-слов
df_vectors["description"] = df_vectors["description"].apply(
    lambda tokens: [word.lower() for word in tokens if (word not in string.punctuation and word not in stop_words and word.isalpha())])

# Лемматизация
df_vectors["description"] = df_vectors["description"].apply(lambda txt: [morph.parse(word)[0].normal_form for word in txt])
```

Получаем итоговый набор данных.

Index	description
0	человек, который, хотеть, описать, выделяться, характер, обнаруживаться, правило, присутствие, следующий, три, черта, аккуратный, бережливый, упрямый
1	от, добавить, характер, четвёртый, обязательный, черта, склонность, садизм, число, скрытый, латентный, форма
2	аккуратность, обозначать, физический, чистоплотность, также, добросовестность, исполнение, иной, род, мелкий, обязательство, человек, аккуратный, смысл, положиться
3	этот, человек, обладать, морально, нравственный, чистоплотность, просто, патологически, честный
4	такой, ребёнок, ответить, телефон, мама, сказать, дом

Таким же образом выполняем предобработку наборов данных df_vectors_pos и df_vectors_neg.

Построение векторных моделей (TF-IDF, WordToVec).

Модель TF-IDF.

```
vectorizer = TfidfVectorizer(ngram_range=(1, 2), max_features = 300)
df_vectors_ = df_vectors.copy()
df_vectors_["description"] = df_vectors_[ "description"].apply(lambda tokens: ' '.join(tokens))

embedding_train_tfidf = vectorizer.fit_transform(df_vectors_[ "description"].values)
```

Модель WordToVec, обученная на описании векторов.

```
simple_model = Word2Vec(df_vectors[ "description"].values, vector_size=300, window=5, min_count=5, workers=4)

w2v_simple = dict(zip(simple_model.wv.index_to_key, simple_model.wv.vectors))

tfidfEmbVect_simple = TfidfEmbeddingVectorizer(w2v_simple)

embedding_train_simple = tfidfEmbVect_simple.fit(df_vectors[ "description"].values).transform(df_vectors[ "description"].values)

y_train_simple = df_vectors[ "vectorId"].values.copy().astype(np.int8)
```

Класс для вычисления среднего вектора с весами TF-IDF набора слов.

```
class TfidfEmbeddingVectorizer(object):
    def __init__(self, word2vec):
        self.word2vec = word2vec
        self.word2weight = None
        self.dim = len(word2vec.popitem()[1])

    def fit(self, X, y=None):
        tfidf = TfidfVectorizer(analyzer=lambda x: x)
        tfidf.fit(X)
        max_idf = max(tfidf.idf_)
        self.word2weight = defaultdict(
            lambda: max_idf,
            [(w, tfidf.idf_[i]) for w, i in tfidf.vocabulary_.items()])

        return self

    def transform(self, X):
        return np.array([
            np.mean([self.word2vec[w] * self.word2weight[w]
                    for w in words if w in self.word2vec] or
                    [np.zeros(self.dim)], axis=0)
            for words in X
        ])
```

Модель WordToVec, предобученная с сайта “rusvectores.org” (корпус “Тайга”).

```
model = KeyedVectors.load_word2vec_format('186/model.bin', binary=True) # tayga-func_upos_skipgram_300_5_2019

w2v = dict(zip(words, model.vectors))

tfidfEmbVect = TfidfEmbeddingVectorizer(w2v)

embedding_train = tfidfEmbVect.fit(df_vectors[ "description"].values).transform(df_vectors[ "description"].values)

y_train = df_vectors[ "vectorId"].values.copy().astype(np.int8)

embedding_train_pos = tfidfEmbVect.transform(df_vectors_pos[ "description"].values)
y_train_pos = df_vectors_pos[ "vectorId"].values.copy().astype(np.int8)

embedding_train_neg = tfidfEmbVect.transform(df_vectors_neg[ "description"].values)
y_train_neg = df_vectors_neg[ "vectorId"].values.copy().astype(np.int8)
```

Построим графики отображений векторов слов в двумерное пространство.

```
tsne = TSNE(n_components=2, learning_rate='auto', init='random')
embedding_new = tsne.fit_transform(embedding_train)

plt.scatter(embedding_new[:, 0], embedding_new[:, 1], c=y_train, cmap='viridis', label=y_train)
plt.legend()
plt.show()
```

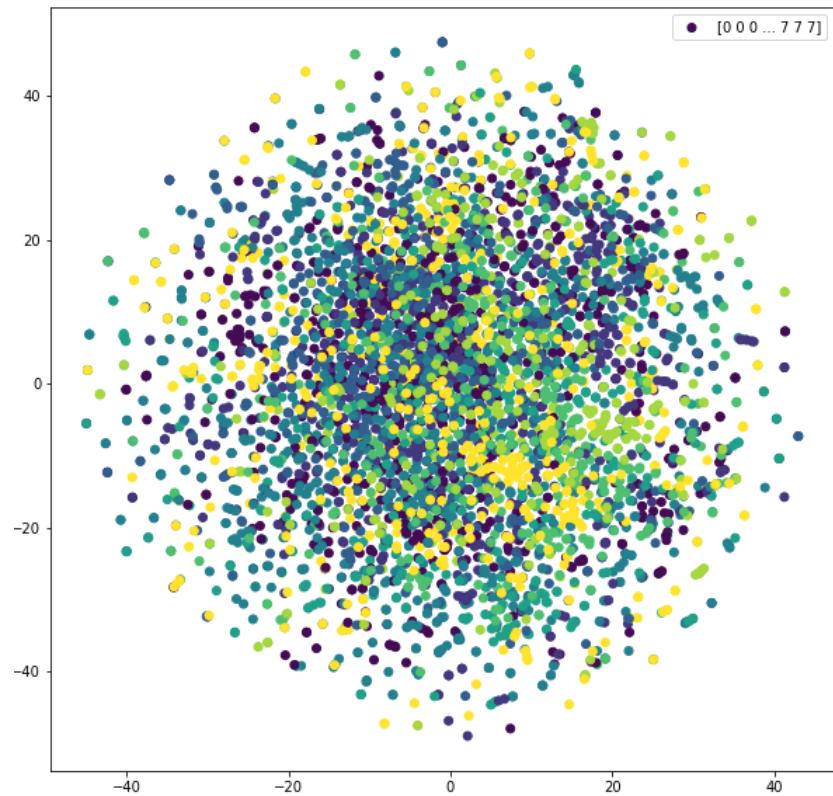


Рис. 1. График отображения средних значений векторов слов из полного описания векторов личности.

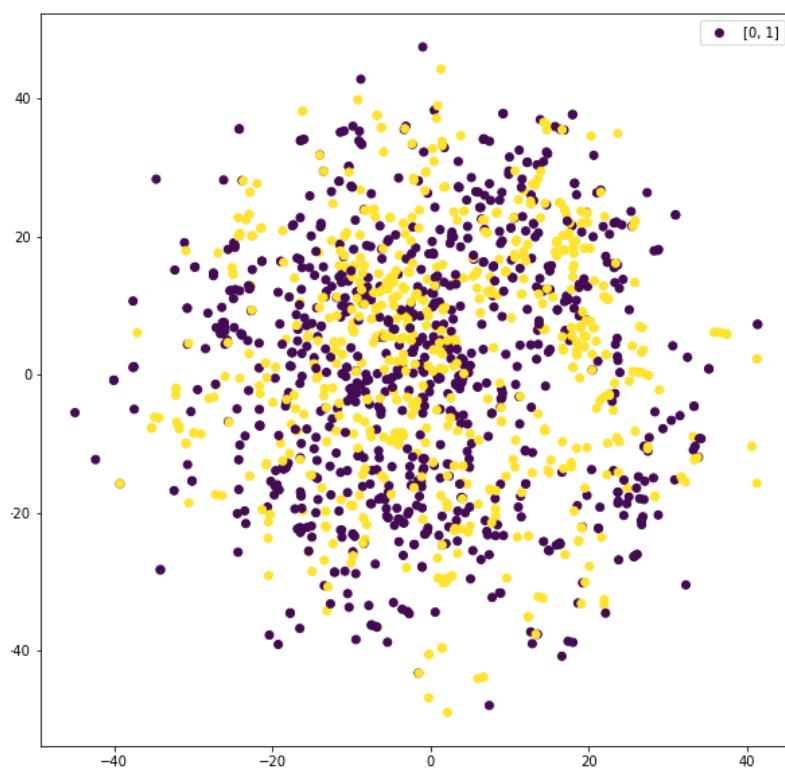


Рис. 2. График отображения средних значений векторов слов из описания коричневого и черного векторов.

На рис. 2 видно, что средние вектора слов коричневого и черного векторов личности нелинейно разделимы.

Построение моделей определения значений векторов личности.

Для определения значений векторов личности по различным текстовым описаниям необходимо построить модель многоклассовой классификации, которая сможет разделять данные на 8 классов и предсказывать вероятность принадлежности к тому или иному классу.

Для построения требуемой модели используем программную библиотеку, разработанную компанией “Яндекс”, CatBoost.

Модель многоклассовой классификации CatBoostClassifier, обученная на значениях векторов слов TF-IDF.

```
classifier_tfidf = CatBoostClassifier(loss_function="MultiClass", iterations=1000, depth=4,
                                         eval_metric="AUC", custom_metric="F1",
                                         random_seed=42, logging_level="Silent", use_best_model=False)#
                                         #, task_type="GPU", de

classifier_tfidf.fit(embedding_train_tfidf, y_train_simple)
<catboost.core.CatBoostClassifier at 0x7f8ee56bfd0>

classifier_tfidf.best_score_

{'learn': {'F1:class=2': 0.6014877789585548,
           'F1:class=0': 0.5778920308483291,
           'F1:class=3': 0.5556021821233739,
           'F1:class=7': 0.6229166666666667,
           'F1:class=4': 0.4676258992805756,
           'F1:class=6': 0.5689948892674618,
           'MultiClass': 1.321765943803622,
           'F1:class=5': 0.5537555228276878,
           'F1:class=1': 0.6363636363636364})}
```

Модель многоклассовой классификации CatBoostClassifier, обученная на значениях векторов слов WordToVec (модель обучена на описании векторов личности).

```
classifier_simple = CatBoostClassifier(loss_function="MultiClass", iterations=4000, depth=4,
                                         eval_metric="AUC", custom_metric="F1",
                                         random_seed=42, logging_level="Silent", use_best_model=False)#
                                         #, task_type="GPU", de

classifier_simple.fit(embedding_train_simple, y_train_simple)
<catboost.core.CatBoostClassifier at 0x7f8f00d759d0>

classifier_simple.best_score_

{'learn': {'F1:class=2': 0.8978873239436621,
           'F1:class=0': 0.8794901506373117,
           'F1:class=3': 0.9075244112578977,
           'F1:class=7': 0.91635687732342,
           'F1:class=4': 0.8725212464589236,
           'F1:class=6': 0.9034013605442177,
           'MultiClass': 0.9601445915992609,
           'F1:class=5': 0.8933002481389578,
           'F1:class=1': 0.9069572506286673})}
```

Модель многоклассовой классификации CatBoostClassifier, обученная на значениях векторов слов WordToVec (предобученная модель с сайта “rusvectores.org”).

```
classifier = CatBoostClassifier(loss_function="MultiClass", iterations=4000, depth=5,
                                 eval_metric="AUC", custom_metric="F1",
                                 random_seed=42, logging_level="Silent", use_best_model=False) #, task_type="GPU", de
classifier.fit(embedding_train, y_train)
<catboost.core.CatBoostClassifier at 0x7f8f00d46dc0>

classifier.best_score_
{'learn': {'F1:class=2': 1.0,
           'F1:class=0': 1.0,
           'F1:class=3': 1.0,
           'F1:class=7': 1.0,
           'F1:class=4': 1.0,
           'F1:class=6': 1.0,
           'MultiClass': 0.14487408896096224,
           'F1:class=5': 1.0,
           'F1:class=1': 1.0}}
```

Гиперпараметры моделей depth и iterations найдены следующим образом:

```
classifier = CatBoostClassifier(loss_function="MultiClass", eval_metric="AUC", custom_metric="F1",
                                 random_seed=42, logging_level="Silent", use_best_model=False) #, task_type="GPU", devices
gs = GridSearchCV(estimator=classifier,
                  param_grid=[{"iterations": [100, 300, 500, 700, 1000, 1500, 2000, 4000, 5000],
                               "depth": [3, 4, 5]}],
                  scoring="accuracy",
                  verbose=4,
                  cv=5)
gs.fit(embedding_train_simple, y_train_simple)
```

Построение моделей определения позитивности векторов личности.

Для определения позитивности векторов личности по различным текстовым описаниям необходимо построить регрессионные модели для каждого вектора личности, где негативному описанию вектора соответствует значение 0.0, а позитивному – значение 1.0.

Для построения требуемой модели также используем программную библиотеку CatBoost.

Модель регрессии CatBoostRegressor, обученная на значениях векторов слов WordToVec (предобученная модель с сайта “rusvectores.org”).

```
regressors = []
for vectorId in range(len(person_vectors)):
    idxs1 = np.where(y_train_pos == vectorId)[0]
    idxs2 = np.where(y_train_neg == vectorId)[0]

    embedding_train_ = np.concatenate( (embedding_train_pos[idxs1] , embedding_train_neg[idxs2]), axis=0)
    y_train_ = np.concatenate( (np.full(len(idxs1), 1.0), np.full(len(idxs2), 0.0)) )

    regressor = CatBoostRegressor(loss_function="RMSE", iterations=2000, depth=5,
                                  eval_metric="AUC", custom_metric="MAE",
                                  random_seed=42, logging_level="Silent", use_best_model=False) #, task_type="GPU"
    regressor.fit(embedding_train_, y_train_)

    print(regressor.best_score_)

    regressors.append(regressor)

{'learn': {'MAE': 0.007583602116345892, 'RMSE': 0.03626679954477956}}
{'learn': {'MAE': 0.004441322808559554, 'RMSE': 0.00642462734770803}}
{'learn': {'MAE': 0.0015232835982804012, 'RMSE': 0.0021234632143844723}}
{'learn': {'MAE': 0.005193722171152738, 'RMSE': 0.007632184324035602}}
{'learn': {'MAE': 0.0006254330149373877, 'RMSE': 0.0008684236024965613}}
{'learn': {'MAE': 0.0016114704174127247, 'RMSE': 0.00224739798627513}}
{'learn': {'MAE': 0.0010382561136828236, 'RMSE': 0.0014176611864518784}}
{'learn': {'MAE': 0.0023008211788608196, 'RMSE': 0.003232922865016173}}
```

Гиперпараметры моделей depth и iterations также найдены с помощью GridSearchCV (для всех моделей одинаковые значения).

Выбор лучшей модели классификации для определения значений векторов личности.

Загрузка и предобработка тестовых данных.

```
df_text = pd.DataFrame([], columns=["description"])

with open("test/Дарвин.txt", encoding="utf8") as rf:
    for sentence in split_text(rf.read().strip()):
        if len(sentence) <= 2: continue
        df_text = df_text.append({"description": sentence}, ignore_index=True)

# Токенизация
df_text["description"] = df_text["description"].apply(lambda txt: word_tokenizer.tokenize(txt))

# Удаление стоп-слов
df_text["description"] = df_text["description"].apply(
    lambda tokens: [word.lower() for word in tokens if (word not in string.punctuation and word not in stop_words and word.isalpha())])

# Лемматизация
df_text["description"] = df_text["description"].apply(lambda txt: [morph.parse(word)[0].normal_form for word in txt])

df_text.head()
```

	description
0	[чарлз, дарвин, воспоминание, о, развитие, м...
1	[я, приходиться, слышать, отец, мнение, челове...
2	[не, обстоять, дело, я, ибо, самый, ранний, мо...
3	[мой, мать, умереть, июль, г, немногое, восемь...
4	[думать, это, забвение, мой, воспоминание, воз...

Проверка качества построенных моделей.

Проверка моделей проводится по книге Ч. Дарвина “Воспоминания о развитии моего ума и характера” и книге Н.В. Гоголя “Нос” на основании следующих утверждений из книги В.К. Толкачева:

1. Яркий представитель коричневого типа — Чарлз Дарвин, об этом говорят не только результаты его литературного творчества, способность собирать и систематизировать научные факты, подробно описывать наблюдаемые явления, но также дневник и факты из работы «Воспоминания о развитии моего ума и характера».

2. Обостренное ощущение носа полностью вылилось в рассказе «Нос», психоаналитическом гимне Гоголя собственной эрогенной зоне. Можно допустить, что Гоголь «видел» ноздрями. Естественно, что, впадая в обонятельный невроз, он с ужасом обнаружил свою нечувствительность к запахам, которую воспринял как потерю носа и красочно перенес ее на майора Ковалева.

Модель многоклассовой классификации CatBoostClassifier, обученная на значениях векторов слов TF-IDF.

```
df_text_ = df_text.copy()
df_text_[ "description" ] = df_text_[ "description" ].apply( lambda tokens: ' '.join(tokens) )

embedding_test_tfidf = vectorizer.fit_transform(df_text_[ "description" ].values)

predict_proba = classifier_tfidf.predict_proba(embedding_test_tfidf)

for idx, vector in enumerate(np.mean(predict_proba, axis=0)):
    print("{:.3f}".format(vector), "-", person_vectors[idx])
```

книга Ч. Дарвина

0.158 – brown

0.130 – black

0.108 – red

0.225 – orange

0.079 – yellow

0.092 – green
0.091 – blue
0.116 – purple

книга Н.В. Гоголя

0.166 – brown
0.128 – black
0.168 – red
0.180 – orange
0.082 – yellow
0.081 – green
0.076 – blue
0.120 – purple

Модель многоклассовой классификации CatBoostClassifier, обученная на значениях векторов слов WordToVec (модель обучена на описании векторов личности).

```
embedding_test_simple = tfidfEmbVect_simple.transform(df_text["description"].values)

predict_proba = classifier_simple.predict_proba(embedding_test_simple)

for idx, vector in enumerate(np.mean(predict_proba, axis=0)):
    print("{:.3f}".format(vector), "-", person_vectors[idx])
```

книга Ч. Дарвина

0.179 – brown
0.137 – black
0.132 – red
0.181 – orange
0.081 – yellow
0.087 – green
0.078 – blue
0.126 – purple

книга Н.В. Гоголя

0.147 – brown
0.131 – black
0.095 – red
0.187 – orange
0.114 – yellow
0.086 – green
0.077 – blue
0.163 – purple

Модель многоклассовой классификации CatBoostClassifier, обученная на значениях векторов слов WordToVec (предобученная модель с сайта “rusvectores.org”).

```
embedding_test = tfidfEmbVect.transform(df_text["description"].values)

predict_proba = classifier.predict_proba(embedding_test)

for idx, vector in enumerate(np.mean(predict_proba, axis=0)):
    print("{:.3f}".format(vector), "-", person_vectors[idx], "{:.3f}".format(pos))
```

книга Ч. Дарвина

0.274 - brown
0.116 - black
0.155 - red
0.147 - orange
0.053 - yellow

0.092 - green
0.056 - blue
0.107 – purple

книга Н.В. Гоголя

0.178 - brown
0.089 - black
0.095 - red
0.223 - orange
0.091 - yellow
0.097 - green
0.048 - blue
0.179 – purple

Видно, что чем лучше модель, тем больше проявляется предполагаемый вектор.

В нашем случае лучшей является модель, обученная на значениях векторов слов WordToVec (предобученная модель с сайта "rusvectores.org"). Далее будем работать с ней.

Проверка моделей на текстовых данных разных тем.

Скачаем с сайта "labyrinth.ru" описания книг, используя следующие ключевые слова: 'кулинар', 'художник', 'бизнес'.

Всего описаний книг - 829, по темам: 'кулинар' - 235, 'художник' - 250, 'бизнес' - 344.

Переводим текстовые описания книг в вектора личности, используя построенные модели многоклассовой классификации и регрессии на основе значений векторов, вычисленных предобученной моделью WordToVec.

```
arr_vectors = []

for text in texts:
    df_texts = pd.DataFrame([], columns=["description"])

    for sentence in split_text(text.strip()):
        if len(sentence) <= 2: continue
        df_texts.append({"description": sentence}, ignore_index=True)

    # Токенизация
    df_texts["description"] = df_texts["description"].apply(lambda txt: word_tokenizer.tokenize(txt))

    # Удаление стоп-слов
    df_texts["description"] = df_texts["description"].apply(
        lambda tokens: [word.lower() for word in tokens if (word not in string.punctuation and word not in stop_words)])

    # Лемматизация
    df_texts["description"] = df_texts["description"].apply(lambda txt: [morph.parse(word)[0].normal_form for word in txt])

embedding_test = tfidfEmbVect.transform(df_texts["description"].values) # Векторное представление слов

predict_proba = classifier.predict_proba(embedding_test)

vectors = []
for vector in np.mean(predict_proba, axis=0):
    vectors.append(vector) # Вектор личности

for idx in range(len(person_vectors)):
    positive = np.mean(regressors[idx].predict(embedding_test))
    vectors.append(positive) # Позитивность вектора

arr_vectors.append(vectors)

arr_vectors = np.array(arr_vectors)
```

Строка массива векторов личности, соответствующего текстовому описанию книг, имеет вид:

v1 v2 v3 v4 v5 v6 v7 v8 pv1 pv2 pv3 pv4 pv5 pv6 pv7 pv8, где

v1 – значение коричневого вектора (от 0.0 до 1.0),

pv1 – позитивность коричневого вектора (от 0.0 до 1.0).

Используя отображение векторов личности и позитивности векторов в двумерное пространство, посмотрим можно ли разделить данные по темам.

На графиках используется следующее цветовое обозначение:

Темно-фиолетовый цвет – описания книг, отфильтрованные по ключевому слову “кулинар”.

Бирюзовый цвет – описания книг, отфильтрованные по ключевому слову “художник”.

Желтый цвет – описания книг, отфильтрованные по ключевому слову “бизнес”.

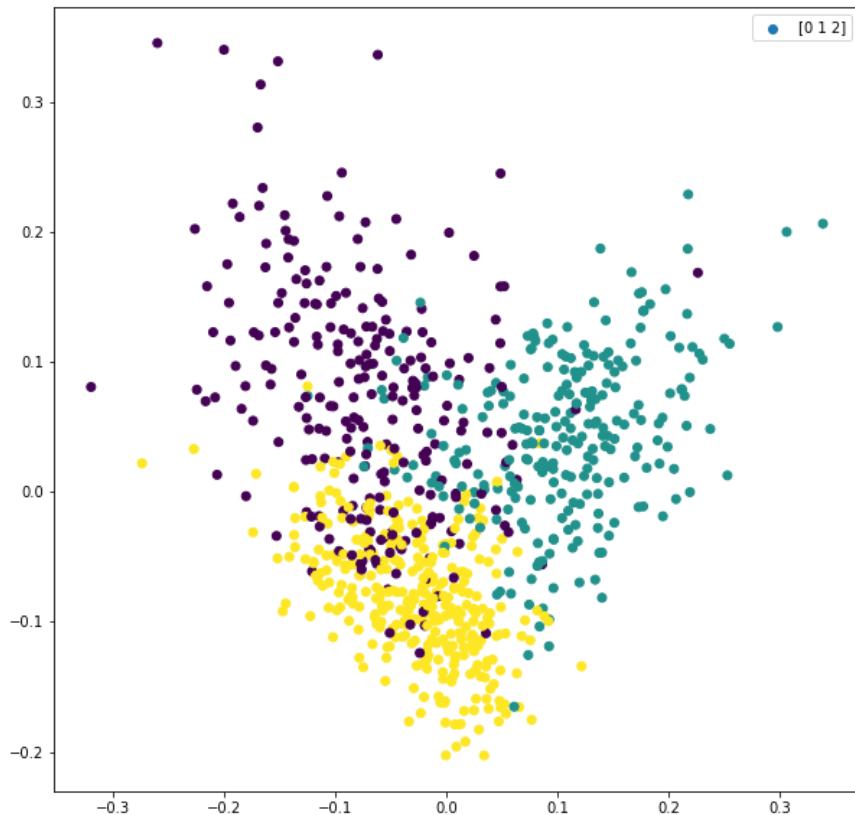


Рис.3. График отображения векторов личности и позитивности векторов описания книг (PCA).

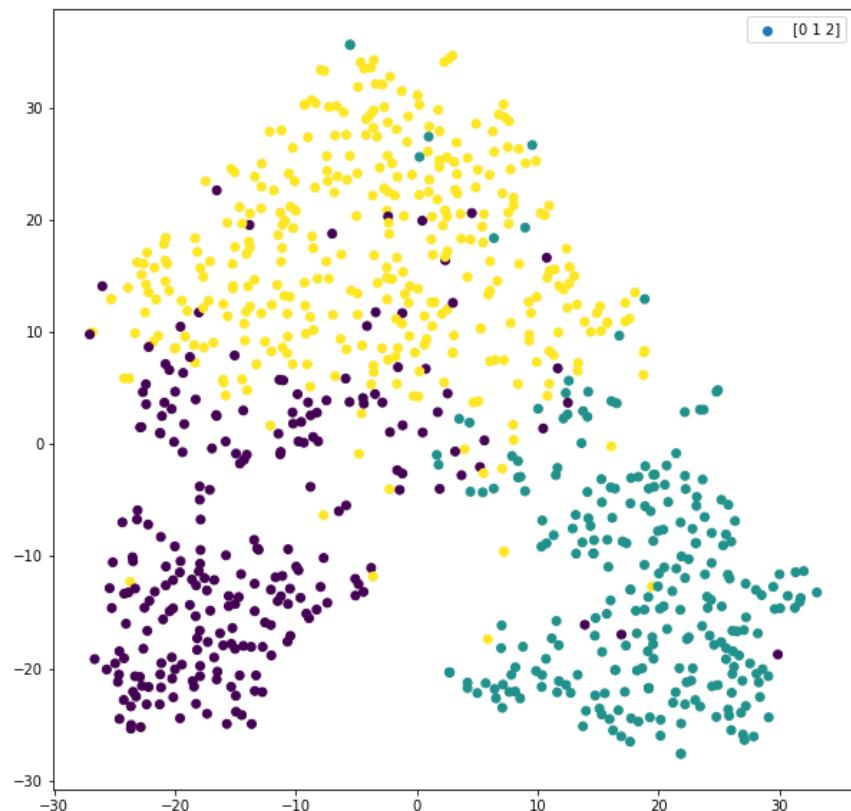


Рис.4. График отображения векторов личности и позитивности векторов описания книг (TSNE).

На рис.4 видно, что данные разделяются на 3 кластера и часть описаний из темы “кулинар” (темно-фиолетовый цвет) располагаются в области описаний темы “бизнес” (желтый цвет), и небольшая часть описаний из темы “художник” (бирюзовый цвет) также располагаются в области описаний темы “бизнес”.

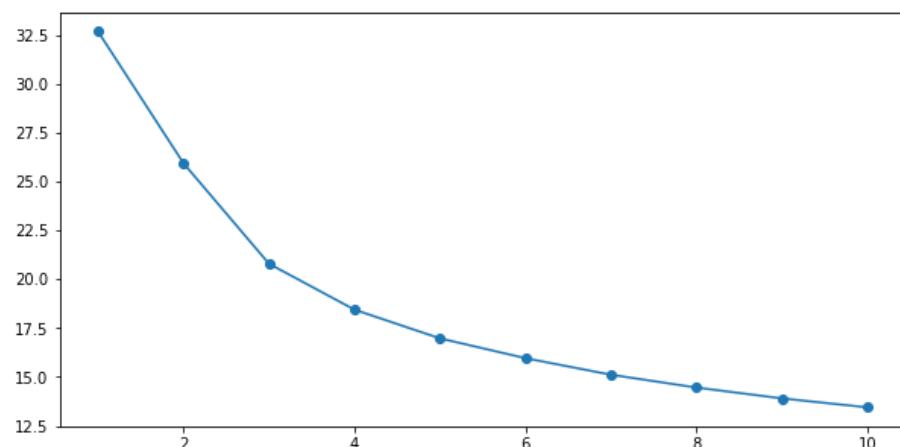


Рис.5. График объясненной вариации в зависимости от количества кластеров.

По графику на рис.5 методом “локтя” также определяется 3 кластера.

Используем TimeSeriesKMeans для кластеризации полученных графиков психологических векторов (вектора личности и позитивность векторов составляют один график).

```
ts_kmeans = TimeSeriesKMeans(n_clusters=n_clusters, n_init=20, max_iter=300, random_state=0, metric="euclidean")
y_pred = ts_kmeans.fit_predict(arr_vectors)
```

По графику силуэта видно, что алгоритм распределил данные на 3 кластера достаточно хорошо.

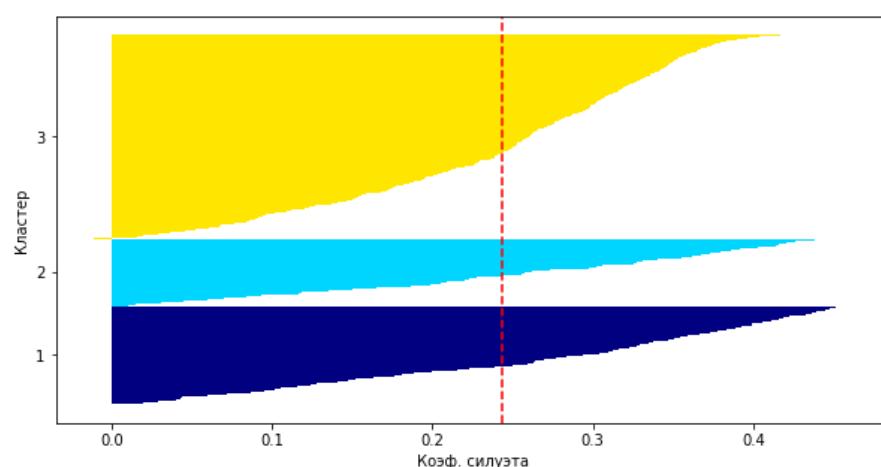


Рис.6. График силуэта при разбиении на 3 кластера.

Алгоритм распределил данные на темы следующим образом:

'кулинар' - 151, 'художник' - 221, 'бизнес' - 457.

Изначальное распределение по темам: 'кулинар' - 235, 'художник' - 250, 'бизнес' - 344.

Посмотрим графики векторов и текстовые описания книг в каждом кластере (выведем 600 символов описания книги).

```
indexes = [1, 0, 2] # Индексы соответствия идентиф. вектора в кластере и идентиф. вектора в исходных данных

labels_ = labels.copy()
for cluster_id in range(n_clusters):
    print('Кластер ' + str(cluster_id + 1))
    fig, (ax1, ax2) = plt.subplots(2, 1, figsize=(10, 7))
    for vectors in arr_vectors[ np.where(y_pred == cluster_id)[0] ]:
        ax1.plot(vectors[:8])
        ax2.plot(vectors[8:])
    plt.show()
print('График векторов личности (вверху), график позитивности векторов (внизу)')

idxs = np.where(labels == indexes[cluster_id])[0]

y_true_cluster = labels[idxs]                      # Истинные метки
y_true_cluster[:] = cluster_id
print('Истинные метки:\n', y_true_cluster)
y_pred_cluster = y_pred[idxs]                      # Предсказанные метки
print('Предсказанные метки:\n', y_pred_cluster)
labels_[idxs] = cluster_id

accuracy = accuracy_score(y_true_cluster, y_pred_cluster)
print('accuracy - ', accuracy, '\n')

df_data = pd.DataFrame(texts[idxs], columns=['text'])
for data in df_data.values[: 7]:
    print(data[0][:200])
    print(data[0][200:400])
    print(data[0][400:600], '\n')

print('Accuracy:', accuracy_score(y_pred, labels_))
```

Кластер 1 (тема 'художник'), преобладают коричневый (0) и зеленый (5) вектора.

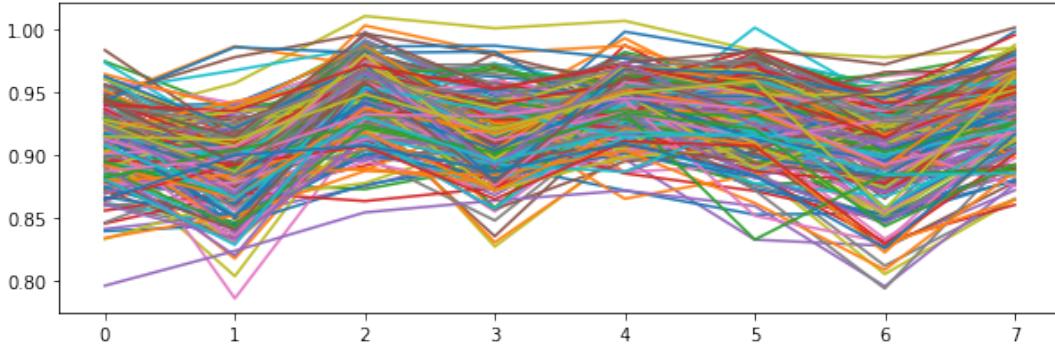
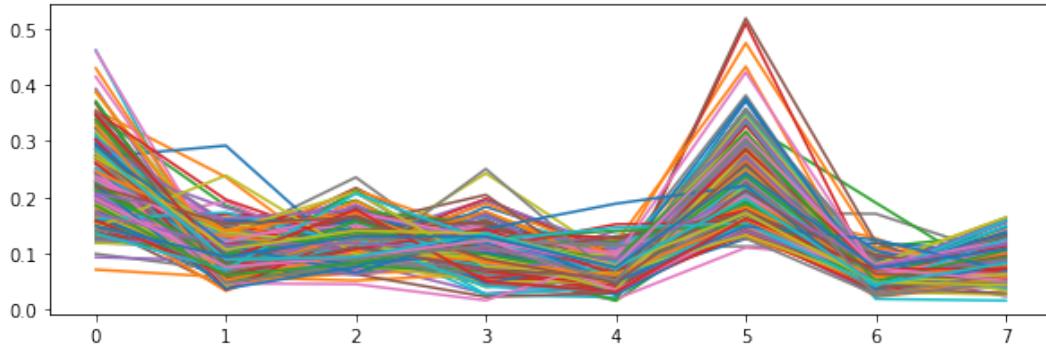


График векторов личности (вверху), график позитивности векторов (внизу)

Истинные метки:

0
0 0]

Предсказанные метки:

[0 2 0 0 0 0 0 0 0 0 2 0 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 2 0 0 0 0 2 0 0 0 0 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0
0 2 0 0 0 0 0 0 0 0 2 0 0 0
0 2 2 0 0 0 0 0 0 0 0 0 2 0 0 2 2 0 0 2 0 0 0 0 0 0 0 2 2 0 0 0 2 2
2 2 2 2 0 0 0 2 0 0 0 0 0 0 0 0 0 0 2 2 0 0 2 0 0 0 0 0 2 0 0 0 0 0
0 2 0 0 2 0 0 0 0 0 0 0 0 2 2 0 2 0 0 0 0 0 0 2 0 2 2 0 2 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 2 2 0 0 0 0 0 0 2 0 0 0 0 0 2]

accuracy - 0.828

"Новая книга лауреата премии им. Бажова, финалиста премий "Большая книга" и "Национальный бестселлер" **О натурщиках и музах известных художников** Анна Матвеева - прозаик, финалист премий "Большая книга", "Национальный бестселлер"; автор книг "Завидное чувство Веры Стениной", "Девять девяностых", "Лолотта и другие парижские истории", "Спрятанные реки" и других. В книге "Картины девушки" Анна Матвеева обращается к судьбам натурщиц и муз известных художников. Кем были женщины, которые смотрят на нас с полотен Боттичелли и Брюллова, Матисса и Дали, Рубенса и Мане? Они жили в разные века, имели раз

Алгоритм распределил в тему “бизнес”.

В настоящей книге собрано более тридцати интервью Энди Уорхола (1962-1987), разделенных по трем периодам: 60-е, 70-е, 80-е. **Изобретательность Уорхола** в интервью обычно оставалась в рамках формата "вопрос-ответ", но намекала, что это лишь формат, и давала интервьюеру право его разрушить. Своими вроде бы банальными ответами художник создавал простор для **творческих способностей интервьюера**, поощряя их еще больше, когда менялся с интервьюером местами или как-то иначе размывал роли интервьюера и интервьюируемого. Можно ли считать эти интервью искусством? Если рассмотреть их в широком контексте, на

Жанр городской монументальной росписи существует на протяжении столетий, даже тысячелетий, но современные граффитисты и мастера уличного искусства (стрит-арта) совсем недавно начали работать в формате XXL. По мере того как росписи увеличивались в размерах, **стрит-арт и граффити сближались**, и сам по себе масштаб отдельных работ стал привлекать внимание и коллег-художников, и широкой публики. Муралы - фасадная живопись огромных размеров - превратились в самостоятельный феномен в искусстве. Книга немецкой художницы с мировым именем Клаудии Вальде посвящена творчеству небольшой группы мастеров либо

Альбом "Достоевский, мировая литература и религия в искусстве художника В. Линицкого" включает работы, написанные в 1950-2000-е годы. **Они воспроизводятся по оригиналам - графическим листам, живописным полотнам**, сохранившимся в государственных и частных архивах. Но ввиду больших утрат оригиналов многие из них даны по фотографиям. В. Линицкий создал портреты Ф.М. Достоевского, иллюстративные серии к его "Бедным людям", "Белым ночам", "Униженным и оскорбленным", "Сну смешного человека", "Братьям Карамазовым"; они появляются впервые в этом издании в исчерпывающей полноте и собраны в особом разделе

Используя особенности графической коллекции Русского музея, визуальный ряд в предлагаемом вниманию читателей издании выстроен с акцентом на **образцы высокого художественного качества рисунка и акварели** в пространстве минувшего века. Избранные в качестве примеров произведения преимущественно принадлежат к разделу станкового творчества, не связанного ни с подготовительной, так называемой лабораторной, графикой, ни с книжной иллюстрацией, ни с театральными постановками. Эти рисунки свидетельствуют не только о смелых и плодотворных решениях профессиональных проблем, о глубине пластического мышления

Ван Гог жил своими картинами, прежде чем их писать. Для создания «Едоков картофеля» он провел три месяца в крестьянской семье, сделал пятьдесят исследований лиц, ел вместе с ними, дышал тем же воздухом. Смотря на картины Ван Гога, мы не ограничиваемся наблюдением за семейной трапезой, мы едим ее вместе с ними. С Ван Гогом живопись становится живой материей. А с Константино д'Орацио жизнь Ван Гога превращается в мастерски сложенный рассказ. Автор

раскроет самые увлекательные секреты, скрытые в галлюцинированных ночных пейзажах и в ярких полях Ван Гога. Мы проникнем среди ветвей оливковых деревьев

Вильнер принадлежит к редкому сословию мастеров, бесконечно преданных однажды избранному пути. Не всё было гладко в его жизни, и долгие годы художник не мог найти то единственное дело, ради которого он пришёл в этот мир. Но удача не обошла его стороной. Сегодня **графическую культуру Петербурга** последней трети XX века невозможно представить без работ Виктора Семёновича, посвятившего своё творчество любимому городу - Петербургу. Начав свой путь в искусстве в 50-х годах прошлого века, он продолжает работать и сегодня. Его литографии представлены во многих музеях нашей страны, в том числе Третьяко

Кластер 1 (тема 'кулинар'), преобладают коричневый (0), оранжевый (3) и желтый (4) вектора.

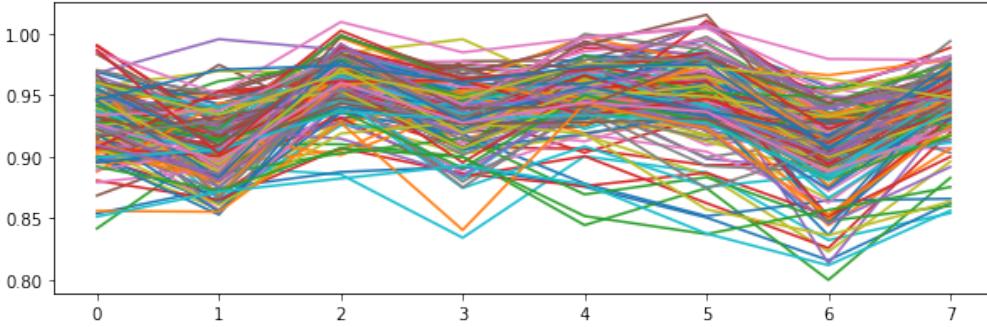
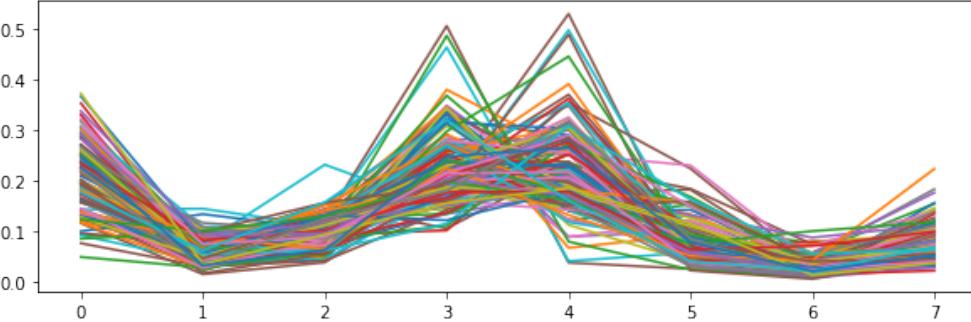


График векторов личности (вверху), график позитивности векторов (внизу)

Истинные метки:

Предсказанные метки:

```
[0 1 2 2 2 1 1 2 2 0 1 1 2 1 2 1 1 0 1 1 1 1 2 1 1 2 2 1 1 1 1 1 1 2 1 1  
1 1 1 2 1 1 2 1 1 2 2 2 2 1 2 1 2 1 1 1 1 1 1 1 1 2 0 1 1 1 2 1 1 2 2 1 1  
1 2 1 1 1 2 1 1 2 2 1 2 1 1 1 2 2 2 2 1 2 1 2 2 2 1 1 1 1 2 1 1 2 1 2 1 2  
1 1 1 2 2 2 1 2 1 1 1 1 2 1 0 1 1 1 2 1 2 1 1 2 1 1 1 1 2 1 1 2 2 1 2 1 1  
1 1 1 1 2 1 2 1 1 1 2 2 1 1 1 1 1 2 1 1 2 1 1 2 1 2 2 1 2 2 2 1 1 1 1 2  
1 2 1 1 1 1 1 1 2 2 2 1 2 1 0 1 0 2 1 2 0 1 1 1 1 2 1 1 1 1 1 1 1 1 0 0 2  
1 2 1 1 1 2 1 1 1 2 1 2 1]
```

accuracy - 0.625

Алгоритм распределил в тему “художник”.

Победитель премии IBPA Benjamin Franklin Awards 2020. В этой **увлекательной книге** от кулинарного историка Энни Грей традиции «Аббатства Даунтон» **представлены** в более чем 100 рецептах. Это рецепты блюд, которые **демонстрируют** обычай семьи Кроули, а также типичны для британской аристократии и прислуги в начале XX века. Книга состоит из двух частей: **роскошные** блюда наверху и простые приемы пищи внизу. Некоторые из рецептов адаптированы для

современности, а некоторые оставлены в первоначальном виде, в котором они были упомянуты или **воссозданы на экране**. Также Энни Грей дает богатую и захватывающую и

"Джейми и друзья" - новая книжная серия, основанная на рецептах, отобранных для читателей всемирно известным британским шеф-поваром Джейми Оливером. Телеведущий, сторонник здорового питания, ресторатор и общественный деятель, Джейми Оливер заслужил доверие, уважение и любовь миллионов профессионалов и начинающих **любителей готовить всех возрастов**. Каждая книга содержит не только рецепты, выбранные самим Джейми, а также полезную информацию по приготовлению, сведения о пищевой и энергетической ценности блюд, удобный рубрикатор и индекс, позволяющий быстро найти нужную информацию в книге. А красоч

Алгоритм распределил в тему "бизнес".

"Диета 80/10/10" доктора Дугласа Грэма - наиболее полный и компетентный труд по сыроедению и фрукторианству, а также здоровому образу жизни и здоровому питанию в целом из всего, что было издано в мире на сегодняшний день. **Изменившая жизни и вернувшая здоровье тысячам людей** - это настольная книга для всех, кто стремится к идеальному здоровью, оптимальному весу, отличной спортивной форме, радостной и наполненной энергией жизни, для тех, кто устал от полумер и стремится только к самому лучшему результату! "По моему глубокому убеждению, из всех существующих сегодня диет наиболее здоровая - это так

Алгоритм распределил в тему "бизнес".

Человек находится в постоянном поиске новых продуктов, инновационных методик, чудодейственных БАДов, которые помогли бы улучшить здоровье, расстаться с лишними килограммами, замедлить старение. Любопытно, что мы часто возвращаемся к тому, с чего все начиналось, к самым основам, традиционной культуре питания и образу жизни, которые складывались веками, и убеждаемся в том, что это, возможно, и есть тот утраченный компонент здоровья, которого нам так не хватает в современном мире. Рафинированные продукты, содержащие синтетические добавки, пестициды, генномодифицированные и фактически утратившие п

Алгоритм распределил в тему "бизнес".

Автор бестселлера "Еда и мозг" раскрывает мощное влияние кишечных бактерий на состояние вашего мозга и предлагает практическую **пошаговую программу для улучшения экологии** вашего кишечника. Согласно результатам последних исследований, которые проводились в наиболее авторитетных научных заведениях по всему миру, здоровье головного мозга и развитие различных его заболеваний в очень сильной степени зависит от того, что происходит в кишечнике человека. Вы не ослышались: состояние вашего кишечника определяет риск возникновения у вас самых разных неврологических расстройств. Пищеварительная система самы

Американский криминально-драматический сериал канала НВО "Клан Сoprano" (The Sopranos), созданный Дэвидом Чейзом, давно вошел в сокровищницу мирового кинематографа и стал культовым. **Познать народ - это познать то, что они едят**, и наоборот, мы представляем вашему вниманию щедрую порцию итальянского достояния, неподвластного времени - The Sopranos Family Cookbook. Nuovo Vesuvio. "Семейный" ресторан, переосмысленный и утонченный. Место, где подают непревзойденную неаполитанскую кухню, является лучшим секретным адресом округа Эссекс. Арти Букко, мастер-шеф этой кухни, лично встречающий каждого гос

Эта книга - сборник чудесных, трогательных зарисовок о волжском детстве. В каждом рассказе спрятан маленький сюрприз - рецепт той далёкой, **вкуснейшей детской еды**, которая, конечно, уже никогда не повторится. Но читаешь - и подмывает попробовать! Охота и леденец на палочке сделать, и уху готовить, и иван-чая хочется набрать и залить кипятком. Налить в кружку, и читать сборник дальше, прихлебывая чай и замирая от радостного узнавания: ой, и у нас в детстве так же было! Ой, и мы то же самое делали, любили, навсегда запомнили - и постараемся передать это дальше. Маша Рупасова Детство остаётся бол

Кластер 3 (тема 'бизнес'), преобладают коричневый (0), красный (2) и оранжевый (3) вектора.

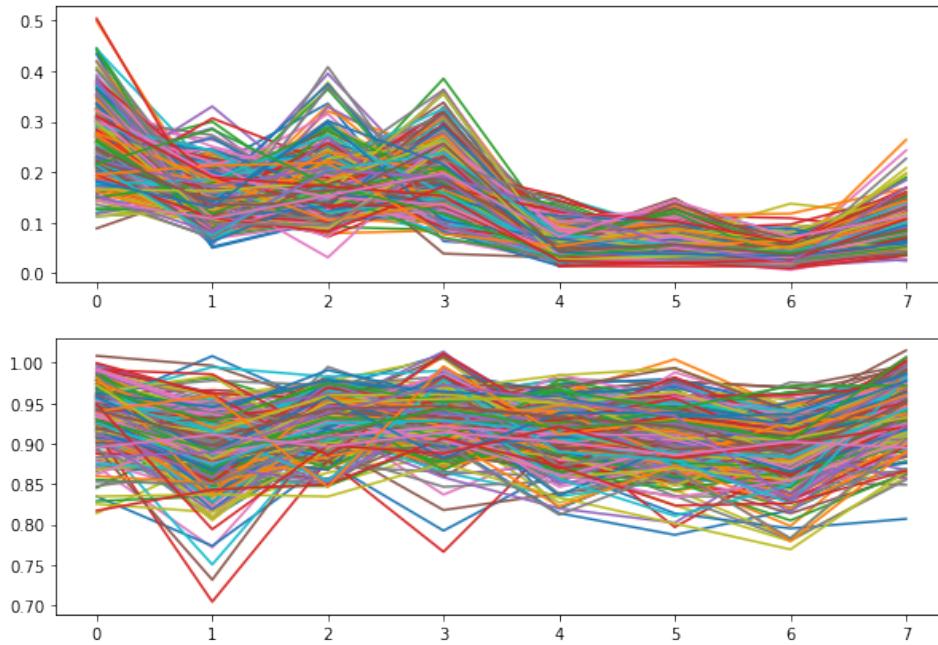


График векторов личности (вверху), график позитивности векторов (внизу)

Истинные метки:

Предсказанные метки:

accuracy - 0.976

Автор написал более 20-ти книг по различным направлениям менеджмента и бизнеса. Все они имеют сугубо практическое, прикладное предназначение. На этот раз автор решил отойти от привычной стези и написал чисто теоретический труд, не имеющий никакого отношения к практическому менеджменту и бизнесу российских компаний. В этой книге речь идет о компаниях, работающих в выдуманной стране (Рашке). Конкретно о системе управленческого учета этих компаний. Выдуманная страна Рашка совершенно не похожа на Россию, где все классно и замечательно, благодаря партии и правительству, уверенно ведущим свой народ

Учебник посвящен аналитическим методам управления бизнес-процессами компании, ориентированным на мировую концепцию устойчивого развития, стейххолдерский подход к формированию и оценке бизнес-модели, а также парадигму отраслевого анализа. Рассмотрены теория и информационная среда бизнес-анализа; методы стратегического анализа и управления рисками; примеры прикладного управленческого инструментария для обоснования стратегии

развития компании в различных секторах экономики: ИТ-секторе, нефтегазовой отрасли, строительном комплексе, транспортно-логистическом секторе, ритейле. В конце каждой главы п

Перед вами книга о базовых **принципах современного торгового бизнеса**. Максим Тверской - предприниматель с тридцатилетним стажем и двадцатилетним опытом работы в ритейле - делится с читателями своими взглядами на торговлю как на единую и целостную систему. Предложенная методология управления торговым предприятием подойдет как для небольшого онлайн- или онлайн-магазина, так и для омниканального оператора или мультиканальной торговой сети на этапе ее перехода к омниканальности. Теоретические основы и практические кейсы готовы к ежедневному применению. Важной частью издания является глава о совреме

Сейчас такое непростое время, что всем хотелось бы каких-то гарантий успешности запуска своего стартапа. Понятно, что никто 100%-й гарантии дать не может, но все же есть один **способ создания бизнеса**, который максимизирует вероятность создания эффективного и успешного бизнеса (правда, при выполнении определенных условий). В наше время нужно выбирать самый надежный способ создания бизнеса. Таким способом является запуск стартапа после прохождения обучения в бизнес-школе для предпринимателей. Естественно, у этого способа, как и у любого другого, есть свои минусы, о которых обязательно нужно знать

Книга посвящена вопросам **построения успешного и устойчивого бизнеса** в любой экономической ситуации и представляет собой одновременно как глубокое исследование, так и сборник практических рекомендаций, следуя которым Вы сможете добиться высочайших результатов в развитии собственного бизнеса или Вашей деловой карьеры, работая наемным менеджером. Как построить эффективную систему управления, как создать успешную команду, как настроить результативные бизнес-процессы, как ставить высокие цели и достигать их минимальными усилиями, как вывести компанию из любого, даже самого глубокого, кризиса на нов

Это 26-я книга Александра Карпова - профессионального эксперта-консультанта по управлению и повышению эффективности компаний (опыт работы - с 1995-го года), президента Ассоциации **эффективного бизнеса**, основателя бизнес-школы "Первая школа предпринимателей". Данная книга выпускается в рамках проекта smart-venture.ru, основной целью которого является содействие всем, кто хочет создавать и развивать свои успешные и эффективные бизнесы. Эту книгу выпустило издательство "Москва", которое на себе опробовало содержащиеся в данной книге методики, технологии и инструменты бизнес-планирования. Использ

Бизнес-ланч. Искусство совместной трапезы и **инструмент успешного бизнеса**. Как выстраивать деловые отношения с помощью ножа и вилки. "Бизнес-ланч - один из лучших способов найти подход к клиенту и установить прочные и взаимовыгодные отношения. Осознав однажды, насколько важно умение выстраивать взаимоотношения с людьми, вы поймете, что совместная трапеза влечет за собой значительный сдвиг в отношениях: "преломляя хлеб" с партнером или коллегой, мы переживаем поистине мистический опыт, который навсегда изменит нашу жизнь. Я научу вас использовать бизнес-ланч как уникальную модель проведения идеа

Accuracy: 0.831

Выводы:

При тестировании построенных языковых моделей получены хорошие результаты в разделении данных по темам (83 %). По графикам векторов видно, что в теме "художник" проявился зеленый вектор, в теме "кулинар" – желтый, а в теме "бизнес" – красный, что полностью соответствует описанию данных векторов в книгах В.К. Толкачева и М. Бородянского. Следует сказать, что тема может быть определена неверно не из-за ошибки модели, а изначально чуть большей принадлежности книги к другой теме, а размещенной в текущей, т.е. 27 % неверно определенных описаний подлежат тщательному анализу.

Кратко из книги В.К. Толкачева:

Зеленый вектор:

Настоящая реальность — это мир яркого света. Отсюда театральность, ярость, демонстративность в поведении визуального типа, некоторый драматизм и острота всех его

переживаний, ибо это «первичные» переживания, связанные с пересечением некоторой новой границы в восприятии. В дальнейшем для этого визуального типа становятся бессознательно желанными просто вспышки света, новое, неожиданное сочетание красок, цветов, оттенков, того, что видится, чем можно наслаждаться вновь и вновь, что придает жизни яркость ощущений.

Красный вектор:

Вечный «искатель приключений и завоеватель». Он раб идеи. Его путь к признанию проложен через осуществление чего-нибудь совершенно грандиозного и уникального, например, через восхождение на самую высокую гору или пересечение океана на плоту без весел и со связанными руками. Такое свершение, по его мнению, затмит всех претендентов на первое место. Он уверен, что всегда добьется аплодисментов, если дольше всех продержится на верхушке флагштока, или станет первым европейцем, который ступил на какую нибудь землю, или пройдет над водопадом по натянутой веревке. Все эти деяния выглядят очень заманчиво, но, конечно, до тех пор, пока они не достигнуты, и искатель приключений вдруг обнаруживает, что покорение очередной вершины ему ничего не дало. Тогда он изобретает себе новый достойный подвиг, который станет пропуском в вечность.

Желтый вектор:

Они — прекрасные кулинары.

Он — ест много, но разнообразно; он все пробует на вкус.

Что может быть приятнее ощущений, исходящих из переполненного вкуснятины желудка и чувства приятной дремы, охватывающей организм? Не в этом ли смысл жизни? Качество жизни наиболее точно измеряется количеством выпитого пива!

Модели определения позитивности векторов, необходимость их использования совместно с моделью определения самих векторов.

Скачаем с сайта “labirint.ru” описания книг, используя следующие ключевые слова: 'ужас', 'любовь'. Всего описаний книг - 448, по темам: 'ужас' - 254, 'любовь' – 194.

Построим графики отображения векторов личности и позитивности векторов в двумерное пространство.

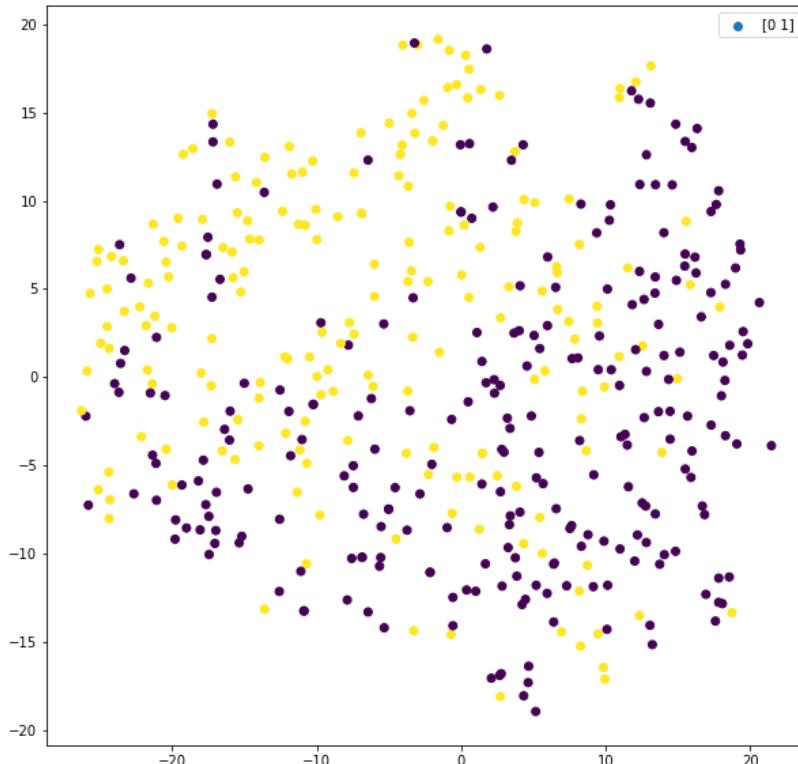


Рис.7. График отображения только векторов личности (TSNE).

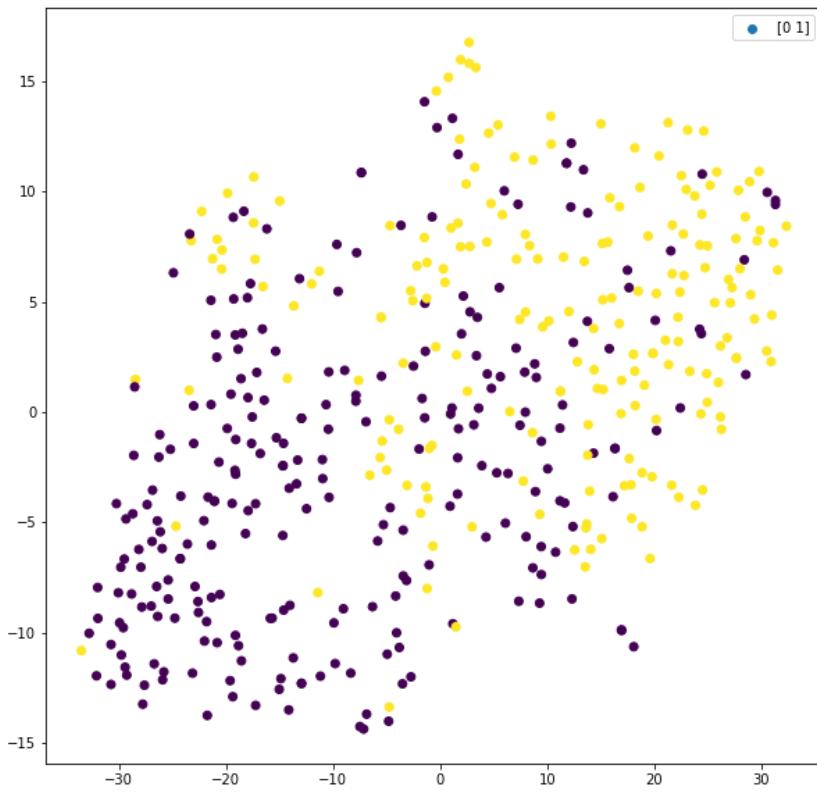


Рис.8. График отображения только позитивности векторов (TSNE).

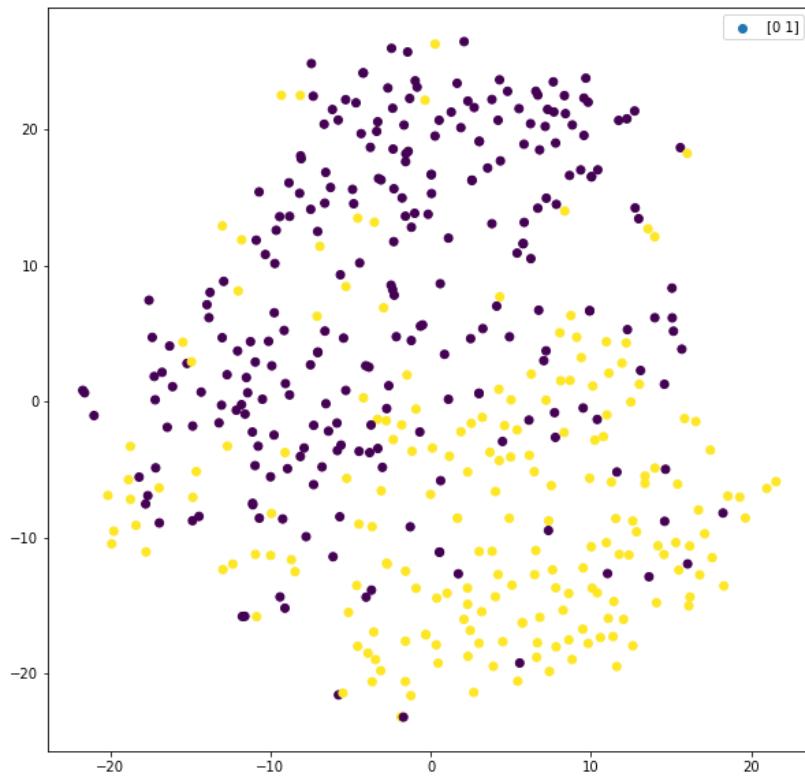


Рис.9. График отображения векторов личности и позитивности векторов (TSNE).

На представленных графиках видно, что данные разделяются лучше (рис. 9), если совместно с векторами личности (рис. 7) использовать и их позитивность (рис. 8).

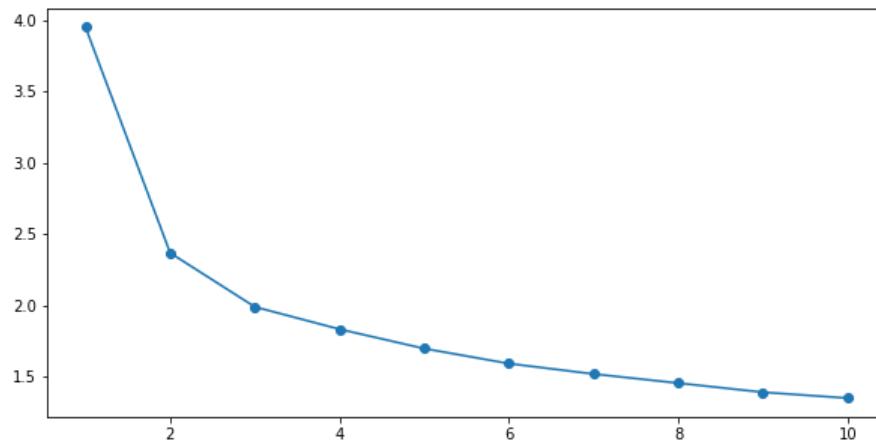


Рис.10. График объясненной вариации в зависимости от количества кластеров.

Используем TimeSeriesKMeans для кластеризации полученных графиков векторов личности и степени позитивности векторов.

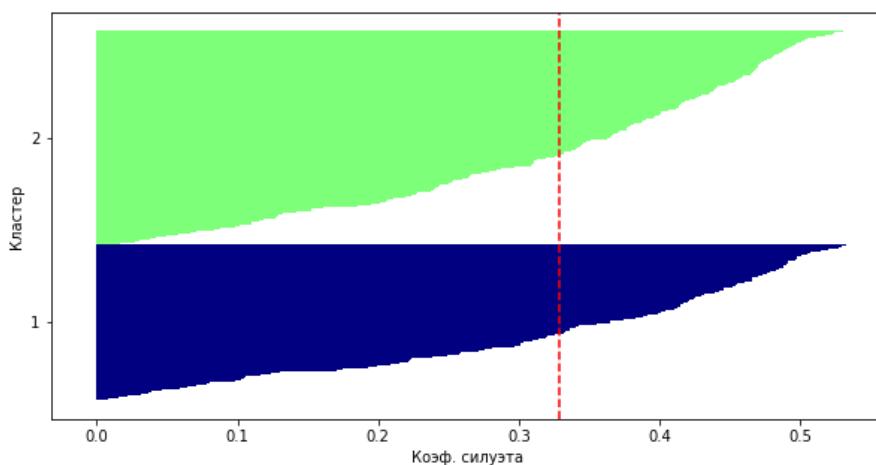


Рис.11. График силуэта при разбиении на 2 кластера.

Посмотрим графики векторов и текстовые описания книг в каждом кластере.

Кластер 1 (тема 'ужас'), преобладают коричневый (0), оранжевый (3) и фиолетовый (7) вектора.

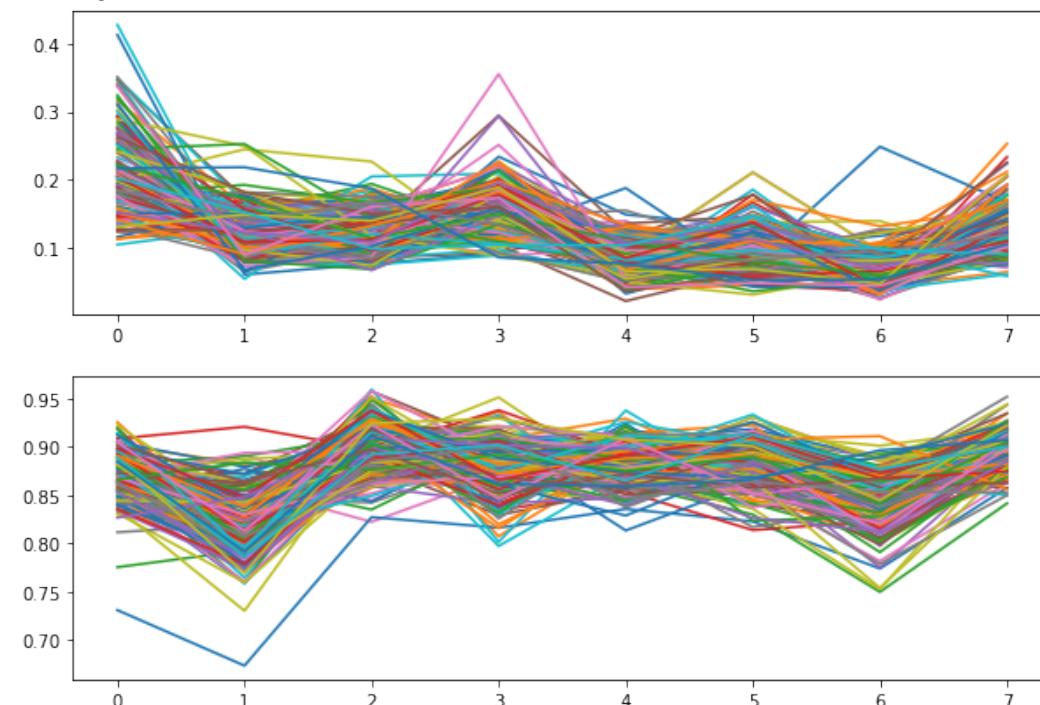


График векторов личности (вверху), график позитивности векторов (внизу)

Истинные метки:

```
[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
```

Предсказанные метки:

```
[0 1 1 1 0 0 0 1 0 0 0 1 1 0 1 0 1 0 1 0 1 1 1 0 0 0 1 1 1 0 1 0 0  
0 1 0 0 0 0 0 1 1 1 0 0 1 0 0 1 0 0 0 1 1 0 0 1 0 0 0 1 1 1 0 0 1  
1 0 0 1 0 1 0 1 0 0 0 0 0 1 0 1 0 0 0 0 0 1 0 1 0 0 1 1 1 0 1 0 1 0  
0 1 0 0 0 0 1 0 1 0 0 0 0 0 1 0 0 1 0 1 0 1 0 0 1 0 1 0 0 0 0 0 0 0  
1 0 0 0 0 0 1 0 0 0 1 1 0 0 1 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 1 1 0 1 0  
0 0 1 0 0 0 1 0 0 1 1 0 1 1 0 0 1 1 1 1 0 1 0 1 0 0 1 0 1 1 1 0 0 1 1  
1 1 1 0 1 0 1 1 0 1 0 0 0 0 1 1 0 0 0 1 0 0 1 1 1 0 0 0 1]
```

accuracy - 0.625

Когда ураган забрасывает циркового фокусника Оскара Диггса из пыльного Канзаса в волшебную страну Оз, **склонный к мошенничеству циркач** полагает, что он поймал удачу за хвост - ведь с помощью своих трюков он с легкостью может добыть в новых землях и славу и богатство. Однако его ждёт встреча с тремя чародейками, Теодорой, Эванорой и Глиндой, которые сомневаются в том, тот ли это великий волшебник, появления которого ждали легковерные жители страны Оз. Волей-неволей, пришелец оказывается втянутым в противостояние эпического масштаба, и ему приходится использовать весь свой "магический" арсенал, с

Алгоритм распределил в тему “любовь”.

Наступил год 1926-й, и мир стоит на краю катастрофы. Где-то за пределами известной реальности пробуждается древнее и невероятно могучее существо, угрожающее стереть человечество с лица Земли. На всех континентах открываются порталы в иные миры, бродят ужасные чудовища и распространяются отвратительные культуры. Лишь горстка отважных сыщиков понимает, что происходит на самом деле. **Эти стойкие защитники человечества должны объединить свои силы в борьбе против Древнего.** Им выпала участь побывать в самых отдалённых уголках земного шара, сразиться с кошмарными тварями и раскрыть жуткие тайны "Древнег

Алгоритм распределил в тему “любовь”.

Семья с двумя детьми живёт на отдалённой ферме. Казалось бы, жизнь этих людей совершенно не отличается от жизни других таких семей, но они живут в месте, которое наполнено ужасными монстрами, реагирующими на любой звук. **Семейство разучило целый комплекс специальных жестов, которые помогают им общаться друг с другом, не издавая ни единого звука.** Кроме того, каждый из членов семьи должен очень тихо передвигаться, чтобы опасные существа их не услышали. Однако дом, где живут дети, не может быть самым тихим местом на земле. Режиссер: Джон Красински Актерский состав: Эмили Блант, Леон Рассом, Джон К

Алгоритм распределил в тему “любовь”.

А. Воронова "Дом тысячи кошек" Однажды, в темном-темном городе посреди вечных дождей, две подруги загадывали желания. **"Хочу влюбиться!" - написала на бумажном самолетике Ника,** а Тишкаписала: "Хочу быть свободной". И самолетики улетели в ночную мглу. Вскоре Ника встретила загадочного мальчика с серебряными глазами, Тишкаполучила способ избавиться от проблем... И только чудо может теперь спасти подруг от внимания страшного существа, готового выполнить их желания. Е.Неволина "Лорд Черного замка" Все! Больше Саша не станет с ними общаться! Ей не нужна ни двоюродная сестра, вечно изображающая из

Вам снова начнут сниться кошмары... **"Паранормальное явление 2"** - такое же поразительное, как и первое. Как только в доме Дэниела и Кристи появился новорожденный ребенок, их начинают терроризировать злые сущности, разбивая вдребезги их прекрасный мир и **превращая его в абсолютный кошмар.** Камеры наблюдения запечатлевают их мучения, делая каждую минуту до

жути реальной. Не исключено, что вам придется рас прощаться со сном - "Паранормальное явление 2" еще долго будет преследовать вас после шокирующего финала. Производство: США, 2010 г. Жанр: ужасы, мистика. Режиссер: Тод "Кип" Уильямс ("Дверь в полу")

Прошло всего четыре месяца после трагической аварии, в которой погибли их родители, и 17-летняя Елена Гилберт и ее 15-летний брат Джереми все еще пытаются оправиться от потери и вернуться к нормальной жизни. Елена всегда была отличной ученицей, красивой и популярной девушкой в школе, но теперь **ей очень тяжело скрывать свою печаль** от внешнего мира. В начале учебного года внимание Елены и ее друзей привлекает новый ученик, загадочный и красивый Стефан Сальваторе. Стефан и Елена тут же чувствуют взаимную симпатию, но Елена даже не подозревает, что **Стефан - вампир**, которому уже сотни лет, и которы

Вам знаком, пробирающий до костей страх? Ужас, который возникает от встречи с непонятным, невероятным... сверхъестественным? Герои историй, вошедших в этот сборник, без раздумий ответили бы "да". Ведь каждый из них столкнулся с тем, что человеческий разум объяснить не в силах. Ожившие скелеты, встающие из могил. Жуткий призрак, бегущий по ночной степи. Бесплотный голос, который вдруг обращается прямо к тебе. Тени с горящими глазами, похищающие энергию душ... Каким бы привычным и безопасным ни казался нам мир, зло не дремлет! И старые тайны ждут тех, кто сумеет их разгадать. А значит, должны найтис

Кластер 2 (тема 'любовь'), преобладают коричневый (0), красный (2), оранжевый (3) и зеленый (5) вектора.

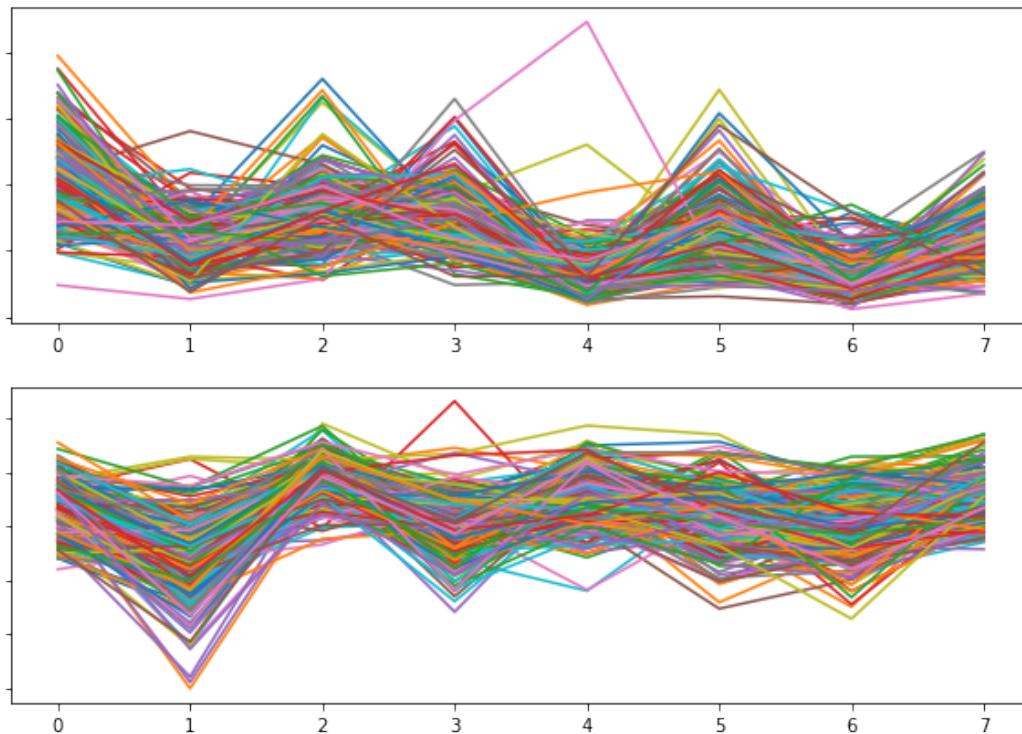


График векторов личности (вверху), график позитивности векторов (внизу)

Истинные метки:

Предсказанные метки:

1 0 1 1 0 1 0 1 1]

accuracy - 0.835

Тик Нат Хан - вьетнамский буддистский мастер. В шестнадцать лет он стал буддистским монахом и духовным искателем. Он пережил преследования, три войны, свыше тридцати лет прожил в изгнании. Тик Нат Хан - настоятель древнего вьетнамского храма, написавший больше ста книг: стихов, рассказов и философских сочинений. Тик Нат Хан основал несколько университетов и общественных организаций. На мирных переговорах в Париже он возглавлял вьетнамскую буддистскую делегацию. Тик Нат Хан номинирован на соискание Нобелевской премии. У него два постоянных места жительства: во Франции и в американском штате Вер

В этой книге классика австрийской детской литературы Кристине Нёстлингер вас ждёт новая порция рассказов про Франца. Францу восемь с половиной лет, и пока ему знакома только взаимная любовь: Франц любит папу, маму, бабушку и брата Йозефа, а те, конечно же, любят его. Но вот Йозеф безответно влюбляется в девочку, с которой едва знаком. Франц за Йозефа очень переживает и решает помочь брату завоевать её сердце. Но стоит ли вмешиваться в чужие отношения? Особенно когда толком не понимаешь, что творится в голове у мальчиков и девочек тридцати лет. Да и к чему вмешиваться, если у самого дружба ра

Авторы подвергают своих героев "испытаниям любовью", ставят на пути любящих сердец препятствия, которые надо преодолеть. Но самые трудные препятствия, как правило, скрываются в глубинах душ самих любящих, и именно их победить тяжелее всего. Золотой век русской литературы не имеет четких временных границ. Он "растянулся" на весь XIX век, от первых произведений сентиментализма и романтизма до триумфа реализма. В этот период были написаны гениальные стихи и поэмы, великие романы и повести. Это и "Бедная Лиза", и "Евгений Онегин", и "Анна Каренина"... Герои их по-разному проходят испытание любовью,

Воспитание девочки в наши дни - задача очень важная и непростая. Каждый родитель мечтает, чтобы его дочь выросла счастливой и гармоничной женщиной, нашла своё место в этом мире, смогла создать и сохранить свою семью, несла мир и радость окружающим. Чтобы этого достичь, нужно с ранних лет давать девочкам примеры, на которые они могут равняться. Эти примеры мы можем почерпнуть из жизни настоящих, не придуманных героинь. Особенно прекрасные женские образы подарила нам история нашей страны. В России родилось и выросло множество истинно великих женщин - мудрых и справедливых правительниц, заботливы

Лилия Нилова — создала бизнес на \$7 миллионов, бизнес-коуч и автор самого популярного блога о маркетинге в русскоязычном Instagram @popartmarketing. Я хочу честно поговорить с вами: нравится ли вам ваша жизнь? Ваша работа? Ваши отношения? Внешний вид и состояние? Если все ответы «да», то я очень рада за вас. Вы по-настоящему счастливый человек. Моя книга для тех, кто устал от себя и образа жизни, который он ведет. Точнее, который он ВЫБРАЛ. Это вы выбрали настоящую реальность или она с вами случилась? Чувствуете разницу? Я хочу рассказать, как брать ответственность за свои решения, разбираться

Алгоритм распределил в тему “ужас”.

Дуэли. Оружие, мастера, факты/ Дуэли. Честь и любовь. Издание в 2-х томах Дуэль. Оружие, мастера, факты: Книга А. Н. Кулинского посвящена очень интересному аспекту мировой истории - дуэлям и дуэльному оружию. На материале архивных документов - военно-судных дел, а также информационных сообщений иностранной печати впервые представлена внушительная статистика русских и иностранных дуэлей XVII-XX вв., включающая 840 эпизодов. Впервые в отечественной литературе приведены некоторые сравнительные данные о количестве дуэлей, состоявшихся в различных странах. Информация, содержащаяся в книге, позволяе

"Пять песен о нас" - романтическая новелла от одного из известнейших экспертов по отношениям Гэри Чепмена, автора легендарного бестселлера "Пять языков любви". Книга подарит вам не только красивую историю о любви, но и несколько уроков по психологии отношений. Марли и Джейкоб поженились в канун Рождества. Теперь, спустя 20 лет, в сочельник они решили развестись. Марли убеждена, что вышла не за того мужчину, Джейкоб устал от того, что его

контролируют. Они спят в разных комнатах и, кажется, больше не готовы спасать свой брак. Их отношения покрыты снегом и льдом, как дорога, по которой они зимни

Accuracy: 0.716

Ниже приведены два противоположных (добро, зло) текстовых описания, найденных с помощью моделей определения позитивности векторов.

Самое позитивное описание (тема “любовь”):

Новинка! Уникальная книга по флористике, которая учит смотреть на растения совершенно по-новому! Благодаря ей составлять чарующие цветочные композиции не составит большого труда. Флорист Эми Меррик расскажет, как собрать идеальный букет к завтраку или в гости, какие акценты сделают даже самую простую композицию прекрасной, а также зачем стоит смотреть на мир через цветочные очки. Каждый разворот книги сопровожден яркими фотографиями или искусственными иллюстрациями. Роскошное оформление, высокое качество Вдохновляющая книга для начинающих флористов содержит секреты искусства аранжировки цветов: от выбора материалов до создания изящных композиций. Эми Меррик – не просто флорист, но и настоящая художница, которая использует все виды растений, которые можно встретить в собственном саду, в поле и в лесу, чтобы помочь нам увидеть знакомые цветы совершенно по-новому. Она учит наслаждаться неожиданными композициями – будь то солнечный весенний букетик в бумажном кофейном стакане, или шикарная зимняя инсталляция в огромной вазе. Вы научитесь видеть прекрасное в простом – в листьях, ветвях, семенных коробочках и даже увядшем цветке, каждый из которых может рассказать свою необыкновенную историю о том месте и времени, в которых они были найдены. Стать флористом со своим уникальным художественным взглядом может каждый. Эта книга расскажет вам о главном правиле составления цветочных композиций — правил нет. Вас ждет необыкновенное путешествие, в котором вы сможете прогуляться по цветочной ферме у побережья штата Вашингтон или сельским просторам Нью-Гэмпшира, изучить искусство составления икебаны в цветочном магазине в Киото. Вы научитесь составлять изящные композиции как профессиональный флорист и собирать небрежные букеты из полевых растений. Здесь собрана антология цветочного вдохновения, любовное письмо к природе от случайного флориста, для которого любой росток — повод для гениального творчества. Лучшая подарочная книга 2019 года по мнению InStyle, Real Simple, Better Homes & Gardens и Wall Street Journal.

Самое негативное описание (тема “ужас”):

Жестокость. Наркотики. Суицид. Добро пожаловать в мир тюремного доктора! Леденящие кровь, душераздирающие и поучительные - так можно охарактеризовать случаи из жизни и истории лечения пациентов, описанные врачом, считавшим своим профессиональным долгом помочь тем, кто находится под арестом. Все ее пациенты - преступники. Но, несмотря ни на что, для них она по-прежнему остается лечащим врачом. Доктор Аманда Браун лечила заключенных в скандально известных тюрьмах Великобритании - вначале в исправительных учреждениях для несовершеннолетних, затем в пресловутой Уормвуд-Скрабс и впоследствии в крупнейшей европейской тюрьме для женщин Бронзфилд. Она видела все - от грязных протестов до удивительных случаев беременности, от чудовищных нападений на заключенных до безжалостных актов причинения вреда самим себе. В этих откровенных, жизнеутверждающих мемуарах Аманда Браун приводит истории и случаи, повлиявшие на ее карьеру, напоминая о помощи тем, к кому большинство из нас скорее проявило бы равнодушие.

Выводы:

Тестирование построенных языковых моделей на противоположных данных (добро, зло) показало необходимость совместного использования модели определения векторов и моделей определения позитивности векторов, т.к. разная по степени позитивности информация может иметь близкие значения психологических векторов.

Применение правила подобия психологических векторов.

Скачаем с сайта "labirint.ru" описания книг, используя следующие ключевые слова: 'кулинар', 'художник', 'медитация', 'бизнес', 'ужас', 'любовь'.

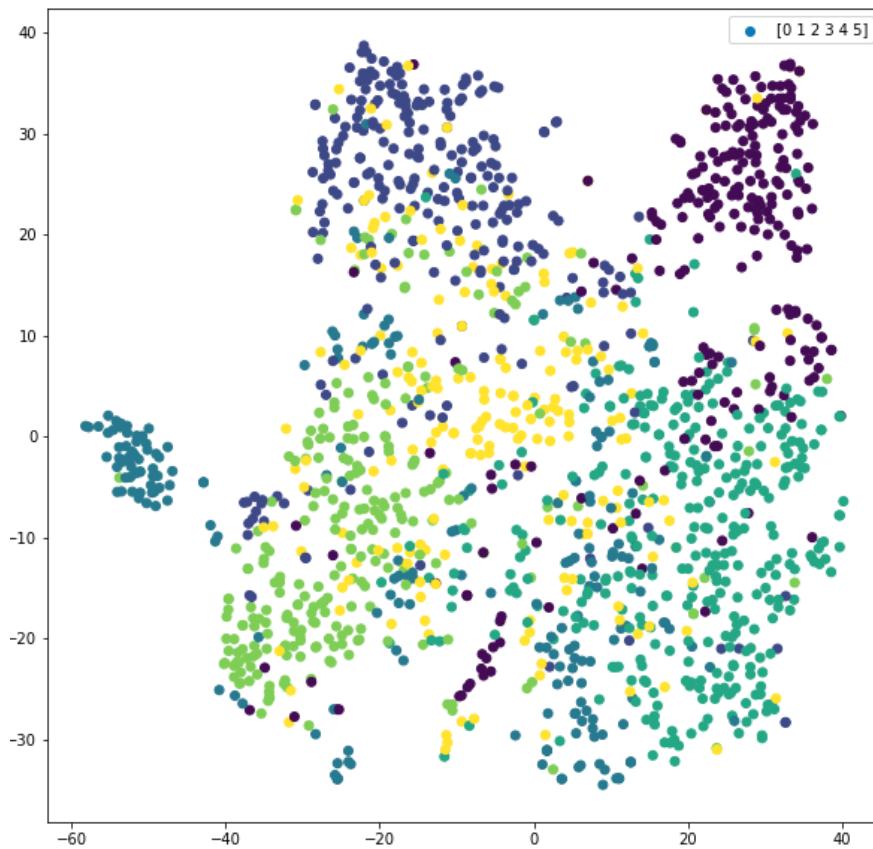


Рис.12. График отображения векторов личности и позитивности векторов (TSNE).

Вектора личности и их позитивность, построенные с помощью языковых моделей по произведению “Ежик в тумане”.

- 0.093 (0.855) - brown
- 0.109 (0.912) - black
- 0.057 (0.901) - red
- 0.200 (0.879) - orange**
- 0.162 (0.894) - yellow**
- 0.122 (0.877) - green**
- 0.097 (0.833) - blue
- 0.160 (0.882) - purple**

По близости векторов (cosine similarity) найдем наиболее подходящее описание книги, которую можно порекомендовать ежику.

Описание книги:

Продолжаете пить каберне, после того как попробовали бутылочку пять лет назад? Наверно и правда, немного боязно уйти в сторону от привычного сорта винограда, но разве вам не хочется попробовать что-нибудь новенькое и интересное? Автор этой книги, Элис Фейринг, утверждает, что о вине можно многое сказать, зная, на какой почве выращивают виноград. В этом путеводителе вина классифицируются не по сортам винограда, не по регионам, не по принадлежности к Старому или Новому Свету, а по горным породам и почвам. Он подробно рассказывает о традициях и техниках виноделия, снимая покров мистики и открывая сомелье и просто любителям совершенно новый взгляд на то, как получать большее удовольствие от вина. Об авторах Элис Фейринг - журналистка, писательница. В 2013 году журнал "Imbibe Magazine" назвал ее человеком года. С 2004 года ведет блог "Линия Фейринг" (alicefeiring.com). Живет в Нью-Йорке. Паскалин Лепельтье -

одна из лучших профессиональных дегустаторов вин. Родом из французского Анжу, и ее любимый сорт винограда - шенен. Глубоко убеждена, что слово "терруар" - не пустой звук.

Вектора описания книги:

0.138 (0.895) - brown
0.074 (0.921) - black
0.090 (0.914) - red
0.194 (0.884) - orange
0.173 (0.948) - yellow
0.131 (0.922) - green
0.086 (0.845) - blue
0.112 (0.891) - purple

Вектора личности и их позитивность, построенные с помощью языковых моделей по произведению Г.Х. Андерсона “Дюймовочка”.

0.090 (0.876) - brown
0.077 (0.894) - black
0.070 (0.938) - red
0.225 (0.910) - orange
0.092 (0.928) - yellow
0.262 (0.898) - green
0.068 (0.883) - blue
0.115 (0.910) - purple

По близости векторов (cosine similarity) найдем наиболее подходящее описание книги, которую можно порекомендовать Дюймовочке.

Описание книги:

Художник-жанрист Алексей Иванович Корзухин (1835–1894) много и упорно трудился над своими произведениями. В своих картинах он старался запечатлеть повседневную жизнь представителей разных сословий. Художник не пытался обличать социальные пороки. Сочувствие к человеку, его желание увидеть в нем доброе и хорошее в конце концов развели его с передвижниками. Его волновали и радостные, и печальные события в жизни людей, и он обстоятельно и с любовью воссоздавал все, что видел и понимал в их жизни. Работал он и над убранством церквей – писал иконы и фрески. Современник сказал о нем: «Светлое и мирное настроение... всегда было настроением А.И. Корзухина». Список репродукций картин А.И. Корзухина: В. Матэ. Портрет художника А.И. Корзухина В монастырской гостинице. 1882 Девичник. 1889 Бабушка с внучкой. 1879 Возвращение из города. 1870 Возвращение с сельской ярмарки. 1868 Петрушка идет! 1888 Шутка Девушка, заплетающая косу. 1883 Воскресный день. 1884 Крестьянские девочки в лесу. 1878 Надоел. 1886 Возвращение с войны Портрет золотопромышленника и исследователя А.М. Сибирякова Сбор недоимок. 1868 У краюшки хлеба. 1890 Перед исповедью. 1877 Портрет Александра III Поминки на кладбище. 1865 Портрет Д.И. Стакеева Разлука. 1872 Девочка. 1877 Канун Рождества. 1869 В красном углу Портрет А.В. Вышеславцева. 1880 Составитель: Астахов Ю.А.

Вектора описания книги:

0.146 (0.813) - brown
0.084 (0.830) - black
0.084 (0.866) - red
0.195 (0.866) - orange
0.116 (0.898) - yellow
0.207 (0.840) - green
0.057 (0.806) - blue
0.112 (0.859) - purple

Описание книги, которая меньше всего бы понравилась Дюймовочке:

Предлагаем вашему вниманию руководство по ремонту и эксплуатации автомобиля "ГАЗель" с бензиновыми двигателями УМЗ-4216, УМЗ-4216-70, Chrysler 2,4L и дизельным двигателем Cummins ISF2.8. В издании подробно рассмотрено устройство автомобиля, даны рекомендации по эксплуатации и ремонту. Специальный раздел посвящен неисправностям в пути, способам их диагностики и устранения. Все подразделы, в которых описаны обслуживание и ремонт агрегатов и систем, содержат перечни возможных неисправностей и рекомендации по их устранению, а также указания по разборке, сборке, регулировке и ремонту узлов и систем автомобиля с использованием стандартного набора инструментов в условиях гаража...

Заключение.

В данной работе были построены языковые модели и проверены на текстовом содержании некоторых книг и текстовых описаниях достаточно большого количества книг. Результаты тестирования показали, что данные модели могут успешно применяться для решения различных задач в современном мире. Следует отметить, что построенные модели не являются идеальными и современные программные алгоритмы позволяют улучшить качество моделей на 7-15 %.

На графике отображения векторов слов двух психологических векторов (рис. 2) видно, что психологические вектора разделяются (проверено для других психологических векторов), из чего следует, что описание каждого психологического вектора является уникальным.

На графике отображения векторов слов всех психологических векторов (рис. 1) видно, что отображения векторов слов в двумерном пространстве составляют окружность, в трехмерном сферу (был построен 3D график), что является довольно интересным фактом.