|  |  |
| --- | --- |
| Gerb-BMSTU_01 | **Министерство образования и науки Российской Федерации**  **Федеральное государственное бюджетное образовательное учреждение**  **высшего образования**  **«Московский государственный технический университет**  **имени Н.Э. Баумана**  **(национальный исследовательский университет)»**  **(МГТУ им. Н.Э. Баумана)** |

ФАКУЛЬТЕТ «Информатика и системы управления» (ИУ)

КАФЕДРА «Системы обработки информации и управления» (ИУ5)

**РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА**

***К НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ***

***НА ТЕМУ:***

**«Обработка набора данных для системы Service Desk»**

Студент группы ИУ5-33М **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** Д.Н. Богомолов

Руководитель **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** Ю.Е. Гапанюк

*2020 г.*

# Введение

Службы поддержки предоставляют пользователям четкую возможность сообщать о проблемах, запросах и запросах и добиться их признания, классификации, владения и принятия мер. Как эта управляется и поставляется эта практика, может варьироваться от физической команды людей на сменной работе до распределенное групп людей, подключенных виртуально, или автоматизированных технологий и ботов. Функция и значение остаются неизменными независимо от модели.

С увеличением автоматизации и постепенным устранением технического долга основное внимание службы поддержки призвано оказывать поддержку «людям и бизнесу», а не просто устранять технические неисправности. Службы обслуживания все чаще используются для решения различных вопросов организованно и скоординировано, а не просто для исправления поломок. Служба поддержки стала жизненно важной частью любой службы.

Для решения задачи автоматизации системы требуются данные для обучения и проверки модели машинного обучения. Так как наборы данных, в основном, состоят из информации, принадлежащей разным типам данных, а модели работают только с числовым типом, то появляется необходимость предобработки набора данных.

# Методология ITIL3

Методология ITIL версии 3 включает в себя следующие этапы оценки жизненного цикла услуги: стратегия, проектирование, преобразование, эксплуатация, непрерывное улучшение. Стратегия является ключевым элементом жизненного цикла ITIL. Она задает планку, которой должна соответствовать IT-услуга, поставляемая организацией. Стратегия услуг, как и другие этапы жизненного цикла ITIL, состоит из различных процессов, комплексов мероприятий, предназначенных для выполнения конкретной задачи. Каждый процесс имеет входные, выходные данные и результаты для формирования изменений.

Схема жизненного цикла услуги ITIL 3 представлена на рис.1.



Рисунок 1. – Схема жизненного цикла услуги ITIL 3

Стратегическая фаза состоит из трех главных процессов: управление финансами, управление спросом и управление портфелем услуг (SPM).

Фаза проектирования услуг учитывает четыре фактора:

* люди (навыки и компетенции, участвующие в предоставлении услуг);
* продукты (технологии и управление);
* процессы (роли и виды деятельности);
* партнеров (производители, разработчики).

На выходе этот этап предполагает формирование подробной спецификации проекта Service Design Package (SDP).

Фаза преобразования услуг включает управление изменениями, управление сервисными активами и конфигурациями (SACM) и управление знаниями по услугам.

Основное внимание в эксплуатационной фазе уделяется проведению и управлению текущей IT-деятельностью. Главными процессами в этой фазе являются управление событиями, управление инцидентами, управление проблемами, управление доступом и запросами на обслуживание.

Последний этап жизненного цикла услуги — непрерывное улучшение услуг.

# ITIL 4. Система ценностей услуг, практики.

Достоинство ITIL 4 по сравнению с другими версиями — новая станет более обобщенной и всесторонней. В новой версии библиотеки к имеющимся процессам добавили фактические ресурсы, необходимые для достижения целей (например, поставщиков или конкретные должности). Новое расширенное понятие стало носить название «практика».

В основе ITIL 4 лежат основные принципы по управлению (guiding principles), которые впервые излагались в ITIL Practitioner. Сейчас они включены в ITIL 4 Foundation.

**ITIL 4 состоит из двух ключевых компонентов:**

• Модель четырех измерений.

Четырёхмерная модель ITIL 4 определяет четыре аспекта, которые следует учитывать для обеспечения целостного подхода к управлению услугами:

• Система ценностей услуг (SVS).

Система ценностей услуг (SVS) представляет, как все компоненты организации взаимодействуют, чтобы способствовать созданию ценности.

1. ITIL 4 практики управления

ITIL 4 включает 34 практики управления как «наборы организационных ресурсов, предназначенных для выполнения работы или достижения цели». Для каждой практики ITIL 4 предоставляет различные типы руководства, такие как ключевые термины и концепции, факторы успеха, ключевые действия, информационные объекты и т. д. Схема, на которой отражены практики ITIL 4 представлены на рис.2.



Рисунок 2. – Практики ITIL 4

В то время как ITIL 3 определил набор процессов, организованных вокруг жизненного цикла сервиса, ITIL 4 описывает принципы, концепции и практики. Это включает в себя ключевые виды деятельности и основные входные и выходные данные для каждой практики, но не подробные спецификации процесса. Этот отход от предыдущего процессно-ориентированного подхода является фундаментальным изменением в ITIL 4, которое позволяет поставщикам услуг применять более гибкие операционные модели.

Таким образом, ITIL 4 не предписывает процессы. Но организациям по-прежнему необходимо определить свои процессы в качестве ключевого элемента своих операционных моделей

# Система Service desk

Службы поддержки предоставляют пользователям четкую возможность сообщать о проблемах, запросах и запросах и добиться их признания, классификации, владения и принятия мер. Как эта управляется и поставляется эта практика, может варьироваться от физической команды людей на сменной работе до распределенное групп людей, подключенных виртуально, или автоматизированных технологий и ботов. Функция и значение остаются неизменными независимо от модели.

С увеличением автоматизации и постепенным устранением технического долга основное внимание службы поддержки призвано оказывать поддержку «людям и бизнесу», а не просто устранять технические неисправности. Службы обслуживания все чаще используются для решения различных вопросов организованно и скоординировано, а не просто для исправления поломок. Служба поддержки стала жизненно важной частью любой службы.

Ключевой момент, заключается в том, что независимо от того, насколько эффективна служба поддержки и её люди, всегда будут проблемы, которые требуют эскалации и поддержки со стороны других команд. Команды поддержки и разработки должны работать в тесном сотрудничестве со службой поддержки для представления и реализации «совместного» подхода к пользователям и клиентам.

Служба поддержки может не быть высокотехнологичной, хотя некоторые из них таковыми являются. Однако, даже если служба поддержки довольно проста, она все равно играет жизненно важную роль в доставке сервисов, и должна активно поддерживаться. Также важно понимать, что служба поддержки оказывает большое влияние на пользовательский опыт и воспринимается пользователями как поставщик услуг.

Службы поддержки не добавляют ценности просто посредством транзакционных действий, например, регистрации инцидентов, но также за счёт понимания бизнес-контекста этого действия и принятия соответствующих мер.

Благодаря повышенной автоматизации, искусственному интеллекту, автоматизации процессов (RPA) и чат-ботам, службы поддержки переходят на предоставление большего количества самостоятельных журналов и решений инцидентов напрямую через онлайн-порталы и мобильные приложения. Снижается влияние на службы поддержки телефонной связи. Производится меньше низкоуровневой работы и появляется больше возможностей сосредоточиться на отличном уровне предоставления сервисов.

Службы обслуживания предоставляют множество каналов доступа. Они включают:

* телефонные звонки, которые могут включать специализированные технологии, такие как интерактивный голосовой ответ (IVR), конференц-связь, распознавание голоса и др.;
* сервисные порталы и мобильные приложения, поддерживаемые сервисом и запросом каталоги и базы знаний;
* чат, через чат и чат-ботов;
* электронная почта для регистрации и обновления, а также для последующих опросов и подтверждений. На основе технологий искусственного интеллекта и машинного обучения оказывается помощь в обработке неструктурированных электронных писем;
* службы поддержки становятся все более распространенными в некоторых секторах, например, образование, где есть высокие пики активности, требующие физического присутствия;
* текстовые сообщения и сообщения в социальных сетях, которые полезны для уведомлений в случае крупных инцидентов и для связи с конкретными группами заинтересованных сторон, но также может использоваться, чтобы позволить пользователям запрашивать поддержку;
* общественные и корпоративные социальные сети и дискуссионные форумы для связи с поставщиками услуг и для одноранговой поддержки.

Некоторые службы поддержки имеют ограниченное окно поддержки, где доступна услуга покрытия (например, с понедельника по пятницу с 08.00 до 20.00).

В некоторых случаях служба поддержки представляет собой осязаемую команду, работающую в одном месте. Централизованная служба поддержки требует поддерживающих технологий, таких как:

* системы интеллектуальной телефонии, включающие интеграцию компьютерной телефонии, IVR и автоматическое распределение звонков;
* системы документооборота для маршрутизации и эскалации;
* системы управления персоналом и планирования ресурсов;
* база знаний;
* запись разговоров и контроль качества;
* инструменты удаленного доступа;
* панель инструментов и инструменты мониторинга;
* системы управления конфигурацией.

В других случаях виртуальная служба поддержки позволяет агентам работать из нескольких мест, которые географически рассредоточены. Виртуальная служба поддержки требует более сложных поддерживающих технологий, включающих более сложную маршрутизацию и эскалацию; эти решения часто основаны на облаке.



Рисунок 3. – Тепловая карта вклада службы поддержки в деятельность цепочки создания стоимости

Персоналу службы поддержки требуется обучение и компетентность в целом ряде областей таких, как технические и деловые области. В частности, им нужно продемонстрировать отличные навыки обслуживания клиентов, такие как сочувствие, анализ инцидентов и расстановка приоритетов, эффективное общение и эмоциональный интеллект. Ключевой навык – уметь полностью понимать и диагностировать конкретный инцидент с точки зрения приоритета бизнеса, предпринять соответствующие действия для решения этой проблемы, используя имеющиеся навыки, знания, людей и процессы.

На рис. 3 показан вклад службы поддержки в цепочку создания стоимости услуг. при этом практика участвует во всех мероприятиях цепочки добавленной стоимости, кроме плана:

* **Улучшение.** Деятельность службы поддержки постоянно отслеживается и оценивается. Поддерживается постоянное совершенствование, согласование и создание ценности. Отзывы от пользователей собираются службой поддержки для поддержки постоянного улучшения.
* **Взаимодействие.** Служба поддержки является основным каналом тактических и оперативных действий. взаимодействие с пользователями.
* **Дизайн и передача.** Служба поддержки предоставляет канал для общения с пользователями о новых и измененных услугах. Сотрудники службы поддержки участвуют в планировании выпуска, тестирования и ранней поддержки.
* **Получение / создание.** Сотрудники службы поддержки могут участвовать в приобретении компонентов для выполнения запросов на обслуживание и устранения инцидентов.
* **Доставка и поддержка.** Служба поддержки является координационным центром для управления инцидентами и запросов на обслуживание.

# Описание и загрузка набора данных

В качестве набора данных (датасета) был выбран журнал событий процесса управления инцидентами, извлеченный из данных, собранных из системы аудита экземпляра платформы ServiceNowTM, используемой ИТ-компанией.

Инцидент в системе Service desk – событие, способное вызвать прерывание работы или снижение качества поставляемого сервиса.

Журнал событий обогащается данными, загруженными из реляционной базы данных, лежащей в основе соответствующей информационной системы технической поддержки Service desk.

Датасет состоит из 36 колонок:

1. **number**: идентификатор инцидента;
2. **incident** **state** - восемь уровней управления процессом управления инцидентами переходят от открытия до закрытия дела;
3. **active**: логический атрибут, который показывает, активна ли запись или закрыта/отменена;
4. **reassignment\_count**: количество раз, когда в результате инцидента группа или аналитики поддержки менялись;
5. **reopen\_count**: количество раз разрешение инцидента было отклонено вызывающим абонентом.;
6. **sys\_mod\_count**: количество обновлений инцидентов до этого момента;
7. **made\_sla**: логический атрибут, который показывает, является ли данный инцидент превысило целевой показатель ОАС;
8. **caller\_id**: идентификатор затронутого пользователя;
9. **opened\_by**: идентификатор пользователя, сообщившего об инциденте;
10. **opened\_at**: дата и время открытия пользователя инцидента;
11. **sys\_created\_by**: идентификатор пользователя, зарегистрировавшего инцидент;
12. **sys\_created\_at**: дата и время создания системы инцидентов;
13. **sys\_updated\_by**: идентификатор пользователя, который обновил инцидент и сгенерировал текущую запись журнала;
14. **sys\_updated\_at**: дата и время обновления системы инцидентов;
15. **contact\_type**: категориальный атрибут, который показывает, какими средствами было сообщено об инциденте;
16. **location**: идентификатор местоположения затронутого места;
17. **category**: описание первого уровня затронутой службы;
18. **subcategory**: описание затронутой услуги второго уровня (относящееся к описанию первого уровня, т. е. к категории);
19. **u\_symptom**: описание восприятия пользователем доступности сервиса;
20. **cmdb\_ci**: (элемент подтверждения) идентификатор, используемый для сообщения о затронутом элементе (не обязательно);
21. **impact**: описание воздействия, вызванного инцидентом (значения: 1-Высокий; 2-средний; 3-низкий);
22. **urgency**: описание срочности, сообщенной Пользователем для разрешения инцидента (значения: 1-Высокий; 2-средний; 3-низкий);
23. **priority**: рассчитывается системой на основе "воздействия" и "срочности".;
24. **assignment\_group**: идентификатор группы поддержки, ответственной за инцидент;
25. **assigned\_to**: идентификатор пользователя, ответственного за инцидент;
26. **knowledge**: логический атрибут, показывающий, использовался ли документ базы знаний для разрешения инцидента.;
27. **u\_priority\_confirmation**: логический атрибут, показывающий, было ли поле приоритета дважды проверено;
28. **notify**: категориальный атрибут, показывающий, были ли сгенерированы уведомления для инцидента;
29. **problem\_id**: идентификатор проблемы, связанной с инцидентом;
30. **rfc**: (запрос на изменение) идентификатор запроса на изменение связано с инцидентом;
31. **vendor**: идентификатор поставщика, ответственного за инцидент;
32. **caused\_by**: идентификатор RFC, ответственного за инцидент;
33. **close\_code**: идентификатор разрешения инцидента;
34. **resolved\_by**: идентификатор пользователя, разрешившего инцидент;
35. **resolved\_at**: дата и время разрешения инцидента пользователем (зависимая переменная);
36. **closed\_at**: дата и время закрытия инцидента пользователем (зависимая переменная).

Загрузка набора данных:

nRowsRead = 1000

df1 = pd.read\_csv('/content/drive/MyDrive/NIR\_Gap/incident\_event\_log.csv', delimiter=',', nrows = nRowsRead)

df1.dataframeName = 'incident\_event\_log.csv'

nRow, nCol = df1.shape

print(f'There are {nRow} rows and {nCol} columns')

Набор содержит 1000 строк и 36 столбцов. Ниже приведены типы данных для каждой из колонок.

df1.dtypes

number object

incident\_state object

active bool

reassignment\_count int64

reopen\_count int64

sys\_mod\_count int64

made\_sla bool

caller\_id object

opened\_by object

opened\_at object

sys\_created\_by object

sys\_created\_at object

sys\_updated\_by object

sys\_updated\_at object

contact\_type object

location object

category object

subcategory object

u\_symptom object

cmdb\_ci object

impact object

urgency object

priority object

assignment\_group object

assigned\_to object

knowledge bool

u\_priority\_confirmation bool

notify object

problem\_id object

rfc object

vendor object

caused\_by object

closed\_code object

resolved\_by object

resolved\_at object

closed\_at object

dtype: object

# Обработка пропусков, смена типов данных.

Все столбцы не имеют пустых значений, но в некоторых из них пустые значения представлены символом «?». Такого рода пропуски данных присутствуют в следующих столбцах:

sys\_created\_by

u\_symptom

problem\_id

assigned\_to

Было принято решение заменить символ «?» на значение NaN при помощи метода replace.

df2['sys\_created\_by']= df2['sys\_created\_by'].replace("?",np.nan)

df2['u\_symptom']= df2['u\_symptom'].replace("?",np.nan)

df2['problem\_id']= df2['problem\_id'].replace("?",np.nan)

df2['assigned\_to']= df2['assigned\_to'].replace("?",np.nan)

df2['problem\_id']= df2['problem\_id'].replace("?",np.nan)

df2['assigned\_to']= df2['assigned\_to'].replace("?",np.nan)

После замены символа на пустое значение распределение значений в столбцах выглядит следующим образом (рис. 4-5):

y = df2["problem\_id"].value\_counts()

data = [go.Bar(x=y.index, y=y.values)]

layout = go.Layout(

    title='problem\_id',

    autosize=False,

    width=400,

    height=400,

)

fig = go.Figure(data=data, layout=layout)

iplot(fig, filename='basic-bar15')

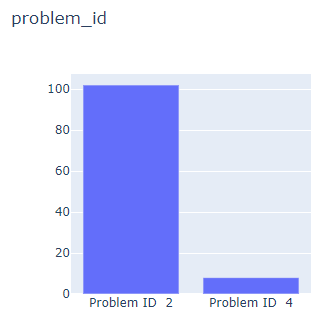


Рисунок 4. – Распределение значений столбца problem\_id

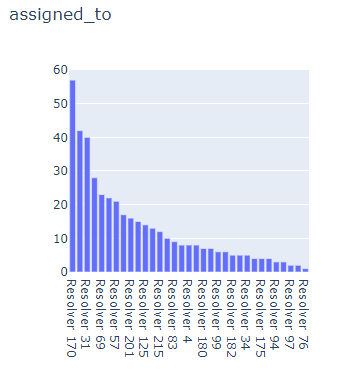


Рисунок 5. – Распределение значений столбца assigned\_to

С помощью метода dropna() строки, содержащие нулевые значения были удалены.

df2= df2.dropna()

Некоторые столбцы датасета содержат строковые значения, пример которых представлен на рис. 6.

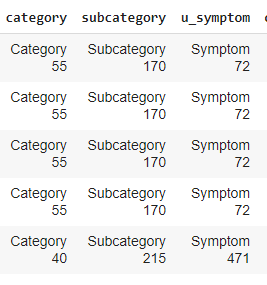


Рисунок 6. – Избыточное представление данных в виде строк

С помощью методов apply() и split() были удалены подстроки, не содержащие полезной информации и тип данных изменён на числовой.

df2['caller\_id'] = df2['caller\_id'].apply(lambda x: int(x.split()[-1]))

df2['opened\_by'] = df2['opened\_by'].apply(lambda x: int(x.split()[-1]))

df2['sys\_created\_by'] = df2['sys\_created\_by'].apply(lambda x: int(x.split()[-1]))

df2['sys\_updated\_by'] = df2['sys\_updated\_by'].apply(lambda x: int(x.split()[-1]))

df2['location'] = df2['location'].apply(lambda x: int(x.split()[-1]))

df2['category'] = df2['category'].apply(lambda x: int(x.split()[-1]))

df2['subcategory'] = df2['subcategory'].apply(lambda x: int(x.split()[-1]))

df2['u\_symptom'] = df2['u\_symptom'].apply(lambda x: int(x.split()[-1]))

df2['assignment\_group'] = df2['assignment\_group'].apply(lambda x: int(x.split()[-1]))

df2['closed\_code'] = df2['closed\_code'].apply(lambda x: int(x.split()[-1]))

df2['resolved\_by'] = df2['resolved\_by'].apply(lambda x: int(x.split()[-1]))

df2['impact'] = df2['impact'].apply(lambda x: int(x.split()[0]))

df2['urgency'] = df2['urgency'].apply(lambda x: int(x.split()[0]))

df2['priority'] = df2['priority'].apply(lambda x: int(x.split()[0]))

Вид ячеек после обработки представлен на рис. 7.

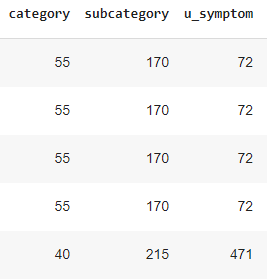


Рисунок 7. – Результат обработки

# Уменьшение количества столбцов

Значения столбца уникального идентификатора number являются служебными, поэтому этот столбец можно удалить.

В столбце contact\_type содержится только одно значение “Phone”. Это можно увидеть при помощи команды value\_counts().

df2.contact\_type.value\_counts()

Phone 1000

Name: contact\_type, dtype: int64

Схожая ситуация в столбцах notify, cmdb\_ci, vendor, caused\_by в которых содержится только по одному значению.

df2.drop(['number', 'contact\_type', 'notify', 'cmdb\_ci', 'vendor', 'caused\_by'], axis='columns', inplace=True)

В столбце rfc 635 ячеек являются пустыми и лишь 8 имеют реальные значения. Так как данных для заполнения пустых значений недостаточно, то было принято решение удалить этот столбец из рассмотрения.

df2.rfc.value\_counts()

? 635

CHG0000127 8

Name: rfc, dtype: int64

df2.drop(['rfc'], axis='columns', inplace=True)

# Работа с категориальными признаками

В датасете в столбце incident\_state присутствуют категориальными признаками. Это можно увидеть на рис.8.

y = df2["incident\_state"].value\_counts()

data = [go.Bar(x=y.index, y=y.values)]

layout = go.Layout(

    title='incident\_state',

    autosize=False,

    width=400,

    height=400,

)

fig = go.Figure(data=data, layout=layout)

iplot(fig, filename='basic-bar15')

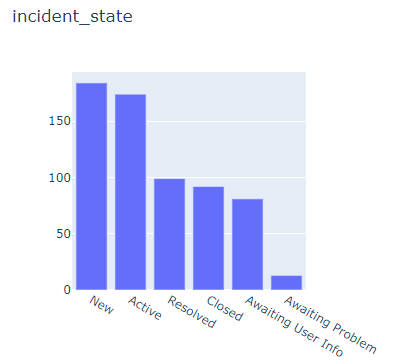


Рисунок 8. – Результат кодирования категориального признака

Для каждой категории был создан отдельный столбец со значениями 0 или 1 при помощи метода get\_dummies(). Затем исходный столбец был удалён. Результат можно наблюдать на рис. 9.

df2 = pd.concat([df2,pd.get\_dummies(df2['incident\_state'], prefix='incident')],axis=1)

df2.drop(['incident\_state'],axis=1, inplace=True)

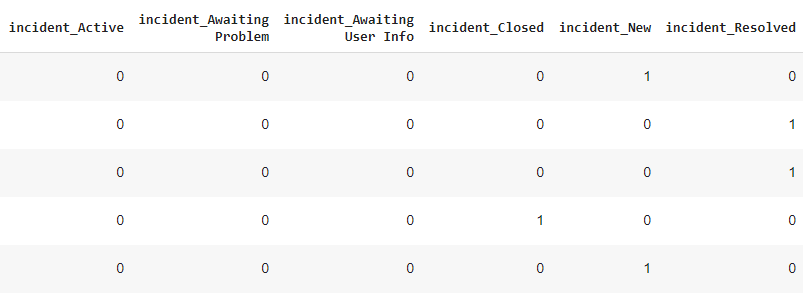


Рисунок 9. – Результат кодирования категориального признака

Результатом проделанной работы является обработанный набор данных, готовый к применению методов машинного обучения.

# Корреляционный анализ

После предобработки данных остались столбцы типа object, в которых содержатся значения в формате дата/время.

Для корреляционного анализа была создана копия набора данных без этих столбцов и построена матрица корреляций.

active int64

reassignment\_count int64

reopen\_count int64

sys\_mod\_count int64

made\_sla int64

caller\_id int64

opened\_by int64

opened\_at object

sys\_created\_by int64

sys\_created\_at object

sys\_updated\_by int64

sys\_updated\_at object

location int64

category int64

subcategory int64

u\_symptom int64

impact int64

urgency int64

priority int64

assignment\_group int64

knowledge int64

u\_priority\_confirmation int64

closed\_code int64

resolved\_by int64

resolved\_at object

closed\_at object

incident\_Active uint8

incident\_Awaiting Problem uint8

incident\_Awaiting User Info uint8

incident\_Closed uint8

incident\_New uint8

incident\_Resolved uint8

dtype: object

df3.drop(['opened\_at', 'sys\_created\_at', 'sys\_updated\_at', 'resolved\_at', 'closed\_at'], axis = 1, inplace = True)

df3.dtypes

active int64

reassignment\_count int64

reopen\_count int64

sys\_mod\_count int64

made\_sla int64

caller\_id int64

opened\_by int64

sys\_created\_by int64

sys\_updated\_by int64

location int64

category int64

subcategory int64

u\_symptom int64

impact int64

urgency int64

priority int64

assignment\_group int64

knowledge int64

u\_priority\_confirmation int64

closed\_code int64

resolved\_by int64

incident\_Active uint8

incident\_Awaiting Problem uint8

incident\_Awaiting User Info uint8

incident\_Closed uint8

incident\_New uint8

incident\_Resolved uint8

Следовательно, набор данных содержит только числовые поля и готов корреляционному анализу.

Матрица корреляций представлена на рис. 10.

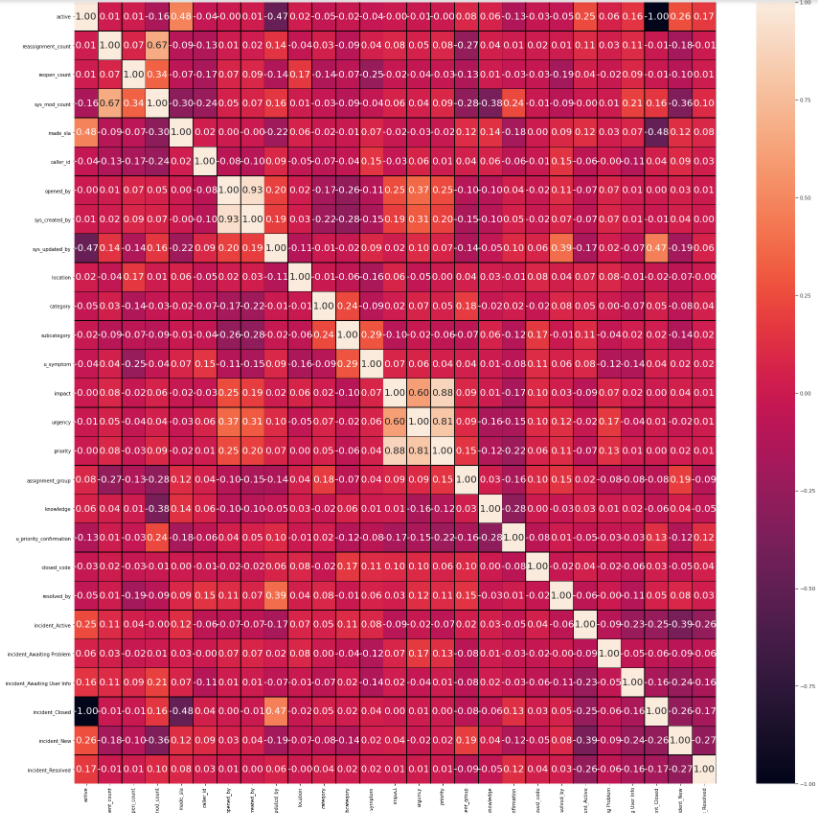


Рисунок 10. – Результат кодирования категориального признака

# Заключение

Service Desk – это мощный инструмент, позволяющий быстрее решать поступающие запросы в соответствии с утвержденным SLA (соглашением об уровнях сервиса), наладить удобное взаимодействие внутри команды техподдержки, а также совместную работу ИТ и разработчиков. Гибкость системы, короткие сроки внедрения и простота встраивания в существующие процессы компании позволят быстро наладить эффективную работу с сервисными запросами. На том или ином этапе жизненного цикла заявки могут пригодиться методы машинного обучения. В основном, это ускорит обработку некорректных заявок и также ускорит процесс классификации заявки и её распределения по отделам технической поддержки. Набор данных, обработанный в рамках исследовательской работы может быть использован для обучения моделей машинного обучения для систем Sercive Desk .

# Список использованных источников

1. Python для анализа данных: обработка данных с помощью Pandas, NumPy и Python. Уэс МакКинни. 2-е издание от 24.10.2017 г.
2. Наука о данных. Базовый курс, Джон Келлехер, Брендан Тирни, 2020 г. – 220с.
3. Репозиторий курса "Методы машинного обучения", магистратура, 2 семестр. URL: <https://github.com/ugapanyuk/ml_course_2020/wiki/COURSE_MMO> (Дата обращения: 20.11.2020)
4. AXELOS: ITIL® Foundation, ITIL 4 Edition. - The Stationery Office; Norwich, UK, February 2019.