

ВАРНЕНСКИ СВОБОДЕН УНИВЕРСИТЕТ
“ЧЕРНОРИЗЕЦ ХРАБЪР”
ФАКУЛТЕТ “СОЦИАЛНИ, СТОПАНСКИ И КОМПЮТЪРНИ НАУКИ”
КАТЕДРА “КОМПЮТЪРНИ НАУКИ”

ХЮСЕИН КАМАЛ ИБРАХИМ
ПРЕДСКАЗВАЩИ БИЗНЕС АНАЛИЗИ С ГОЛЕМИ ДАННИ
ЗА УСЛУГИ ОТ ТЕЛЕКОМУНИКАЦИОННАТА ИНДУСТРИЯ

АВТОРЕФЕРАТ

на дисертационен труд за присъждане
на образователна и научна степен „доктор“,
професионално направление 4.6. "Информатика и компютърни науки",
докторска програма „Информационни системи и технологии,
информатика и компютърни науки“

Научен ръководител:
проф. д-р Владимир Димитров

ВАРНА
2025 г.

ВАРНЕНСКИ СВОБОДЕН УНИВЕРСИТЕТ
“ЧЕРНОРИЗЕЦ ХРАБЪР”
ФАКУЛТЕТ “СОЦИАЛНИ, СТОПАНСКИ И КОМПЮТЪРНИ НАУКИ”
КАТЕДРА “КОМПЮТЪРНИ НАУКИ”

ХЮСЕИН КАМАЛ ИБРАХИМ
ПРЕДСКАЗВАЩИ БИЗНЕС АНАЛИЗИ С ГОЛЕМИ ДАННИ
ЗА УСЛУГИ ОТ ТЕЛЕКОМУНИКАЦИОННАТА ИНДУСТРИЯ

АВТОРЕФЕРАТ

на дисертационен труд за присъждане
на образователна и научна степен „доктор“,
професионално направление 4.6. "Информатика и компютърни науки",
докторска програма „Информационни системи и технологии,
информатика и компютърни науки“

Научен ръководител:
проф. д-р Владимир Димитров

Рецензенти:
доц. д-р Веселина Спасова
доц. д-р Йоаннис Патиас

ВАРНА
2025 г.

Дисертацията е с обем 190 страници. Състои се от въведение, изложение в четири глави, заключение, списък с използвани съкращения, 66 фигури, 19 таблици. Използваната библиография включва 141 литературни източника (книги, статии и доклади от конференции) на английски език. Във връзка с темата на дисертацията са направени четири публикации.

Авторът на дисертацията е докторант в самостоятелно обучение в катедра „Компютърни науки“ към факултет “Социални, стопански и компютърни науки” на Варненския свободен университет „Черноризец Храбър“.

Публичната защита ще се проведе на 11.09.2025 г. от 14,00 ч на открито заседание на научното жури в заседателната зала на ВСУ “Черноризец Храбър“. Материалите по защитата са на разположение в катедра „Компютърни науки“, както и на сайта на университета www.vfu.bg, раздел „Докторанти“.

Въведение

Мотивация

Телекомуникационният сектор е изправен пред нарастване на оплакванията на клиентите поради увеличеното търсене и по-високите очаквания за обслужване. Високите нива на отпадане на клиенти, достигащи до 30% годишно, представляват значителни финансови предизвикателства, като американските и европейските телекомуникационни компании губят 4 милиарда долара всяка година. Задържането на съществуващи клиенти е много по-рентабилно от придобиването на нови. Проактивното управление на оплакванията е от съществено значение, тъй като обратната връзка от клиентите предлага ценна информация за подобряване на услугите и растеж на бизнеса. Компаниите, които третират оплакванията като възможности, могат да повишат удовлетвореността, да намалят отпадането на клиенти и да засилят лоялността към марката. Прогнозните анализи могат да помогнат на телекомуникационните оператори да се справят с проблемите, преди те да ескалират, подобрявайки ефективността и клиентското преживяване.

Описание на проблема

В индустрията за мобилни телекомуникации удовлетвореността на клиентите е от решаващо значение, но много проблеми с услугите остават неотчетени или се появяват косвено чрез канали като социалните медии. Повечето телекомуникационни компании реагират реактивно на оплаквания, често твърде късно, за да предотвратят разочарованието и отпадането на клиентите. Въпреки че телекомуникационните компании събират огромни количества данни, те рядко се използват за прогнозиране на недоволството. Това изследване има за цел да запълни тази празнина, като прилага контролирано машинно обучение, за да идентифицира модели и да прогнозира потенциални оплаквания. Чрез ранно прогнозиране на проблемите, телекомуникационните компании могат да предприемат проактивни действия за подобряване на обслужването, намаляване на отпадането и повишаване на лоялността на клиентите.

Цел и задачи

Това изследване има за цел да разработи подход за прогнозиране на мълчаливите оплаквания на клиентите в телекомуникационните компании. Този подход трябва да използва техники за машинно обучение и да бъде тестван върху реални телекомуникационни данни.

Целта на **това** изследване е процесите на управление на взаимоотношенията с клиенти (CRM).

Предмет на това изследване е отливът на клиенти чрез данните от CRM.

За постигане на тази цел, изследването се фокусира върху следните задачи:

- **Задача 1:** Да се проучат процесите за обслужване на клиенти с цел да се открият свързаните с тях пропуски в интерпретацията и реакцията на оплакванията на клиентите.
- **Задача 2:** Да се проучат съществуващите техники за машинно обучение и тяхната приложимост към целта на изследването.
- **Задача 3:** Да се създаде модел за факторите за оплаквания в телекомуникационна компания въз основа на постигнатите резултати от предишните задачи.
- **Задача 4:** Да се внедри разработеният модел за оплаквания на клиенти.
- **Задача 5:** Да се тества и оцени предложения модел с данните от телекомуникационната компания Lebanon Touch.

Цели на изследването

Това изследване е фокусирано върху подпомагане на мобилните телекомуникационни компании да преминат от реактивен към по-проактивен начин за обработка на оплакванията на клиентите. С помощта на контролирано машинно обучение целта е да се забележат ранни предупреждения, че клиентът може да е недоволен, преди реално да подаде оплакване.

За да се постигне това, изследването се ръководи от следните цели:

1. **Да се проучат често срещаните причини и модели зад оплакванията на клиентите** в индустрията за мобилни телекомуникации, като се използват исторически данни от реални взаимодействия с клиенти и използване на услуги.
2. **За да се съберат и подготвят телекомуникационни набори от данни**, трябва да се изпълнят следните стъпки: почистване на данните, избор на подходящи характеристики и трансформиране на данните във форма, подходяща за машинно обучение.
3. **Да се изградят и обучават отделни модели за машинно обучение с наблюдение**, които могат да се учат от събраните данни, за да предскажат кои клиенти е вероятно скоро да подадат оплаквания.
4. **Да се избере оптимизиран обучен модел за всяка от изследваните техники за машинно обучение**
5. **Да се оцени и сравни производителността на различни алгоритми за машинно обучение**, като се използват показатели като точност, прецизност и изчерпаемост, за да се определи кой модел работи най-добре за този конкретен проблем.
6. **Да се оценят резултатите по начин, който е логичен за телекомуникационните оператори**, като се подчертаят най-важните фактори, които водят до оплаквания, и се предложат анализи, въз основа на които могат да се предприемат действия.

7. **Да се разработи практическа рамка или система**, която телекомуникационните компании биха могли потенциално да възприемат, за да наблюдават и прогнозират недоволството на клиентите, за да им помогнат да подобрят качеството на услугите и клиентското преживяване.

Структура на дисертацията

Ще бъдат следвани следните стъпки, както е посочено по-долу:

Глава 1 разглежда свързаната литература относно поведението на клиентите при оплаквания и анализа им в мобилната телекомуникационна индустрия, използвайки техники за машинно обучение. Също така обсъжда ливанският пазар на мобилни телекомуникации и склонността на клиентите към оплаквания в мобилната индустрия.

Глава 2 прилага методологията на работата, включително описание на експерименталните входни данни, процеса на откриване данни и проектирането на алгоритми за машинно обучение.

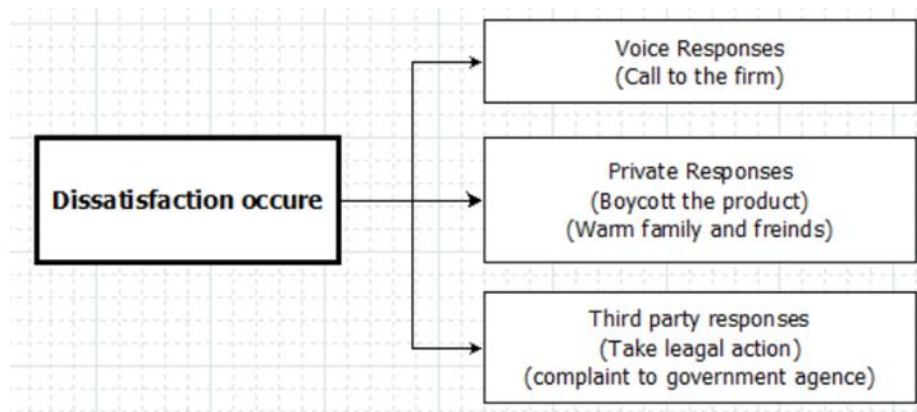
Глава 3 обяснява метода на изследване и как платформата е реализирана, използвайки методи за машинно обучение, базирани на Logistic Regression, Random Forest и ANN Sequential с най-важните активиращи функции, за да се намерят най-добрите и най-точни параметри за прогнозиране на оплакванията на клиентите .

Глава 4 представя експерименталните резултати, обсъжда и подчертава точността на всеки метод и обобщава с резултатите точността на модел, който може да се използва в CRM система. След това се обобщава извършената работа, завършва се с направения принос и се дават някои бележки за бъдеща работа.

Глава 1 Анализ на оплакванията на клиентите в телекомуникационните индустрии

Идентифицирането и разрешаването на скрити оплаквания от клиенти е от решаващо значение, тъй като те не винаги могат да изразят недоволството си, докато не достигнат критична точка. Следователно са необходими проактивни действия, за да се открият евентуални проблеми с качеството на обслужване и да се адресират, преди да се превърнат в оплаквания. Системите за ранно откриване са от съществено значение за този процес, защото позволяват на фирмите да предвиждат промени в моделите и бързо да се справят с новите проблеми, когато възникнат. Тези алгоритми са полезни в много различни области, като например финансова търговия, прогнозиране на продажбите и прогнозиране на отпадането на клиенти. Жалбата на клиент е искане или изискване на клиента за промяна на продукт или услуга.

Освен това, през 1996 г. Сингх разделя потребителските реакции на три групи: директно свързване с бизнеса, за да изразят опасения, въздържане от продукта и предупреждаване на други лица и предприемане на съдебни действия или подаване на жалби до правителството.



Горната фигура категоризира реакциите на клиентите в три общи поведения: директно оплакване до компанията, отказ от услугата и негативна реклама от уста на уста, както и правни/официални действия. Тя обосновава причината, поради която системите за ранно предупреждение в телекомуникациите са важни, за да се избегне нарастването на недоволството.

Таблица 1 предоставя разбивка на причините за оплакванията на клиентите и вероятността им да се оплакват от видовете проблеми с обслужването.

	Coverage Problem	Difficulty Obtaining Information	Incorrect Billing	Incorrect Charges	Breach of Contract	Difficulty Unsubscribing	Delayed Service
Number of users	536	73	37	34	27	13	9
Number of complainers	153	63	35	31	23	12	8
Percentage of complainers (%)	28.5	86.3	94.6	91.2	85.2	92.3	88.9

Note: N = 729 declared dissatisfaction at least once

Таблица 1: Разпределение на оплакванията сред потребителите въз основа на категориите причини

Таблицата очертава факторите, допринасящи за недоволството на клиентите, обхващащи несъответствия в покритието, трудности при достъпа до информация, неточности при фактурирането, неправилни такси, нарушения на договорни споразумения и забавяния в предоставянето на услуги. Основният фокус е върху телекомуникационното покритие, което се отнася до ефективния географски обхват на сигнала на базова станция. Това покритие се влияе от фактори като терен (напр. планини), сгради и технологии (напр. LTE). Свързаността на мобилните устройства се определя от силата на сигнала, излъчван от базова станция.

Проблемът с недостатъчното покритие често е резултат от проблеми с качеството на услугите. Недостатъчното покритие обикновено е в основата на повечето проблеми, свързани с услугите. Следователно основната причина за недоволството на клиентите е проблемът с покритието. Това означава, че подобряването на покритието може да намали недоволството на клиентите. Интересното е, че за разлика от други проблеми като например неправилно фактуриране, които могат да съставляват до 95% от оплакванията, клиентите, които са недоволни от проблеми с покритието, имат по-нисък процент на оплаквания. За разлика от по-ясните проблеми като забавена доставка или неточно фактуриране, клиентите може да се затруднят да предоставят доказателства в подкрепа на оплакванията си относно проблеми с покритието, което обяснява разликата в броя на оплакванията на клиентите.

В ливански контекст, Ливан има един от най-конкурентните мобилни пазари в световен мащаб, известен със своята интензивна конкуренция. Между 2010 и 2018 г. годишният темп на растеж на мобилните клиенти се е увеличил с повече от 13,5%, което показва значително развитие. Следователно мобилните

оператори са се фокусирали силно върху разширяването на пазарните си дялове, като изпреварват конкурентите си в придобиването на абонати.

Конкурентната среда се промени отново, когато проникването на пазара надхвърли 100% през 2012 г., което показва, че мобилните абонати са повече от действителното население на пазара. Това предполага ограничен потенциал за растеж на нови клиенти. В резултат на това процентът на отпадане на абонатите намалява до 1,5%, а темпът на растеж на абонатите пада до по-малко от 1%.

$$\text{Churn rate} = \frac{\text{Number of customers lost in a period}}{\text{Number of customers at the beginning}}$$

1.2 Защо е важно да се предвиждат оплаквания?

С нарастването на пазарния дял, поддържането на висока удовлетвореност на клиентите става по-сложно поради по-широката и по-разнообразна клиентска база. Макар че оплакванията на клиентите често се възприемат негативно, те предлагат ценна информация и възможности за подобрене, ако се управляват добре. Правилното обработване на оплакванията може да повиши лоялността и конкурентоспособността. Някои недоволни клиенти обаче мълчат, което може да доведе до пропуснати възможности. Вместо винаги да се опитват да надминат очакванията, бизнесът трябва да се съсредоточи върху постоянното постигане на реалистични цели и наблюдение на ключови показатели за ефективност (KPI), за да идентифицира проблеми и непрекъснато да подобрява клиентското преживяване.

Многобройни проучвания в сектора на мобилните телекомуникации са изследвали прогнозирането на отпадането на клиенти, използвайки техники за машинно обучение. Отпадането е от решаващо значение за мобилните оператори, тъй като пряко влияе върху тяхната рентабилност. Разширяването на клиентската база на дадена услуга често е свързано с нейната рентабилност. Мозер и др. фокусират своето проучване върху използването на статистически методи за машинно обучение за прогнозиране на отпадането на клиенти, идентифициране на подходящи стимули за подобряване на задържането на клиентите и максимизиране на рентабилността на оператора; изследването на Мозер и др. подчертава колко важно е качеството на разговорите при определяне на удовлетвореността на абонатите.

Factor	Importance	Nature of data required for prediction
Call quality	21%	Network
Pricing options	18%	Marketing, Billing
Corporate capability	17%	Market, Customer service
Corporate service	17%	Customer service
Customer communications	10%	Market, Customer service
Coverage	7%	Network
Handset	4%	Application
Billing	3%	Billing
Roaming cost	3%	Market, Billing

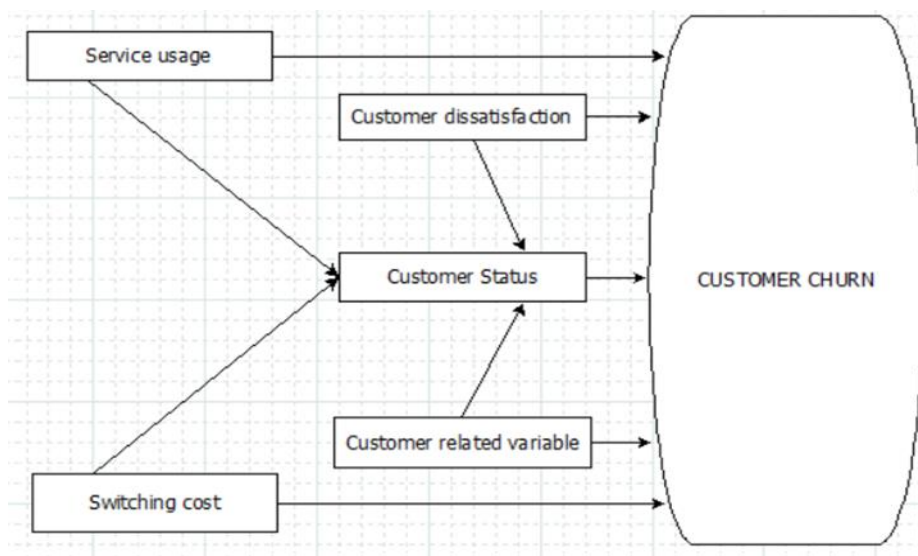
** Source: Mozer, 2000*

Фигура 5 Фактори, влияещи върху удовлетвореността на абонатите [73]

Фигура 5 показва значимите фактори, като например качество на обажданията, точност на фактурирането и достъпност на услугите, които влияят върху удовлетвореността, и които също са променливи, използвани в ML моделирането.

1.3 Рамка за анализ на отпадането на клиенти

„Полезността на една теория зависи от нейната способност да развива знанията чрез репликация, разширяване и обобщение“. След внимателен анализ на литературата за тежестта на клиенти и преглед на настоящите модели, използвани в тази област, за рамка на изследването е избран цялостен модел. Според модела на Ан и др., няколко променливи влияят върху отпадането на клиенти. Очаква се четири ключови компонента да повлияят пряко върху отпадането на клиенти, докато се предполага, че статусът на клиента го медира или косвено влияе.



Фигура 6 Фактори, влияещи върху отпадането на клиенти [77]

Фигура 6 показва значителни фактори за отлив на клиенти, като например недоволство, ниски нива на използване и демографски фактори. Тази фигура директно подпомага обосновката за използване на машинно обучение при прогнозиране на отпадането на клиенти и оплакванията.

Проучванията постоянно показват, че недоволството на клиентите допринася значително за отлива на клиенти в телекомуникационния сектор. Ключови фактори, влияещи върху удовлетвореността, включват качеството на обажданията и мрежата, ценообразуването, точността на фактурирането и обслужването на клиентите. Удовлетвореността се оформя от това доколко услугите отговарят на очакванията на клиентите; като лошото качество на обажданията и неуспехите в обслужването водят до по-ниска удовлетвореност. Данни от телекомуникационните компании, като оплаквания, неуспешни обаждания и продължителност на взаимоотношенията с клиентите, могат да се използват за измерване на недоволството. Изследванията показват, че отливът се увеличава с повече неуспехи в обслужването и оплаквания и намалява с по-дълги взаимоотношения с клиентите. Тези данни помагат за разбирането и прогнозирането на поведението на клиентите.

За това изследване бяха избрани следните променливи: обща сума на сметките, непогасени салда, месечен брой минути ползване, брой повиквания, уникални повиквания и количество кратки съобщения. Въз основа на тези фактори бяха предложени определени теории за нивото на използване на услугите.

Следователно, те предположиха следното:

- 1- Вероятността клиент да напусне е обратнопропорционална на размера на кредита
- 2- Месечните минути за ползване имат отрицателна корелация с вероятността клиент да напусне
- 3- Вероятността клиент да напусне е отрицателно корелирана с броя на обажданията
- 4- Вероятността клиентът да напусне е обратнопропорционална на количеството изпратени кратки съобщения.
- 5- Вероятността клиент да напусне е обратнопропорционална на броя на уникалните обаждания.
- 6- Има връзка между разходите за таксуване и удовлетвореността на клиентите

Смяната на доставчик на услуги често е свързана с различни разходи – като закупуване на нов телефон, загуба на номер или плащане на такси за прекратяване на връзка, които могат да обезкуражат клиентите да напуснат, дори и да са недоволни. Тези разходи за смяна се разделят на три категории:

- Разходи за обучение (усилия за адаптиране към нова услуга),
- Изкуствени/договорни разходи (санкции като такси за предсрочно прекратяване) и
- Транзакционни разходи (финансови разходи за преминаване към нов доставчик).

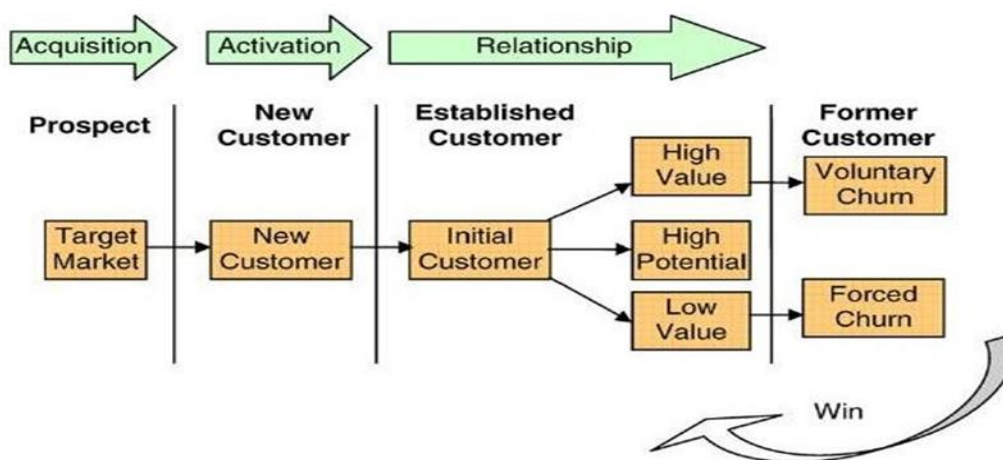
Високите разходи за преминаване към друг доставчик могат да намалят отпадането на клиенти, но ако качеството на услугата е твърде лошо, клиентите все пак могат да изберат да напуснат. Видът услуга и свързаните с нея бариери за преминаване към друг доставчик влияят върху задържането на клиентите.

Минали проучвания са изследвали как демографските данни на клиентите, като възраст, пол, доход и начин на плащане, влияят върху отпадането на клиенти, но резултатите са смесени. Някои изследвания показват, че възрастта и доходът влияят върху поведението при смяна на доставчика, докато други не са открили ясна връзка. По подобен начин влиянието на пола, трудовия статус и характеристиките на домакинствата върху отпадането варира в различните проучвания. Като цяло демографските фактори могат да играят роля в прогнозирането на отпадането, но тяхното влияние не е последователно подкрепено.

1.4 Управление на взаимоотношенията с клиенти и основни понятия

Идеята, че развитието на силни взаимоотношения с клиентите може да доведе до по-голяма лоялност, задържане на клиенти и рентабилност, доведе до създаването на управление на взаимоотношенията с клиентите (CRM) през 90-те години на миналия век. Въпреки широкото му приемане, все още няма универсално признато определение за CRM.

Както е илюстрирано на Фигура 10, жизненият цикъл на клиента предоставя информация за развитието и промените в взаимоотношенията с клиентите във времето:



Фигура 7 Жизнен цикъл на клиента

Фигура 10 визуализира растежа на клиентите от придобиване до потенциален отлив, валидирайки перспективата на CRM и обосновката за използване на модели за прогнозиране, фокусирани върху задържането на клиенти.

1.5 Пример за отлив на клиенти и въздействие върху пазара

Отливът на клиенти се отнася до клиенти, които прекратяват взаимоотношенията си с дадена компания, което се отразява негативно на приходите и стойността на жизнения цикъл на клиента (CLV). Процентът на отлив се увеличава с развитието на пазарите, особено в телекомуникациите, където годишният отлив може да достигне 20 – 40%. Видовете отливи включват доброволен (иницииран от клиента) и принудителен (иницииран от компанията). Разходите за преминаване към друга компания, качеството на услугата и удовлетвореността на клиентите влияят върху поведението на

отлива. Компаниите използват както целенасочени (проактивни/реактивни), така и нецелеви стратегии за управление на отлива. Прогнозните модели са от съществено значение за ранното идентифициране на клиенти в риск. Изследванията показват силна връзка между удовлетвореността, лоялността и задържането на клиенти, като подчертават необходимостта от наблюдение на качеството на услугата и потребителското поведение, за да се намали отливът.

1.6 Заключение

Тази глава очертава основните причини за оплакванията на клиентите в телекомуникациите, като акцентира върху качеството на обслужване и тихия отлив на клиенти. Тя подчертава необходимостта от проактивно управление и дискутира машинното обучение като инструмент за прогнозиране на отлива. Фокусът се измества от придобиването към задържането на клиенти, формирайки основата за следващата глава, посветена на методологията.

Глава 2 Методология

2.1 Въведение

Основата на маркетинговите изследвания е проектното изследване, тъй като той подготвя почвата за събиране и анализ на данни. Тази глава ще разгледа тънкостите на проектното изследване, като ще подчертае неговата ключова роля в изследователския процес.

2.2 Проект на изследването

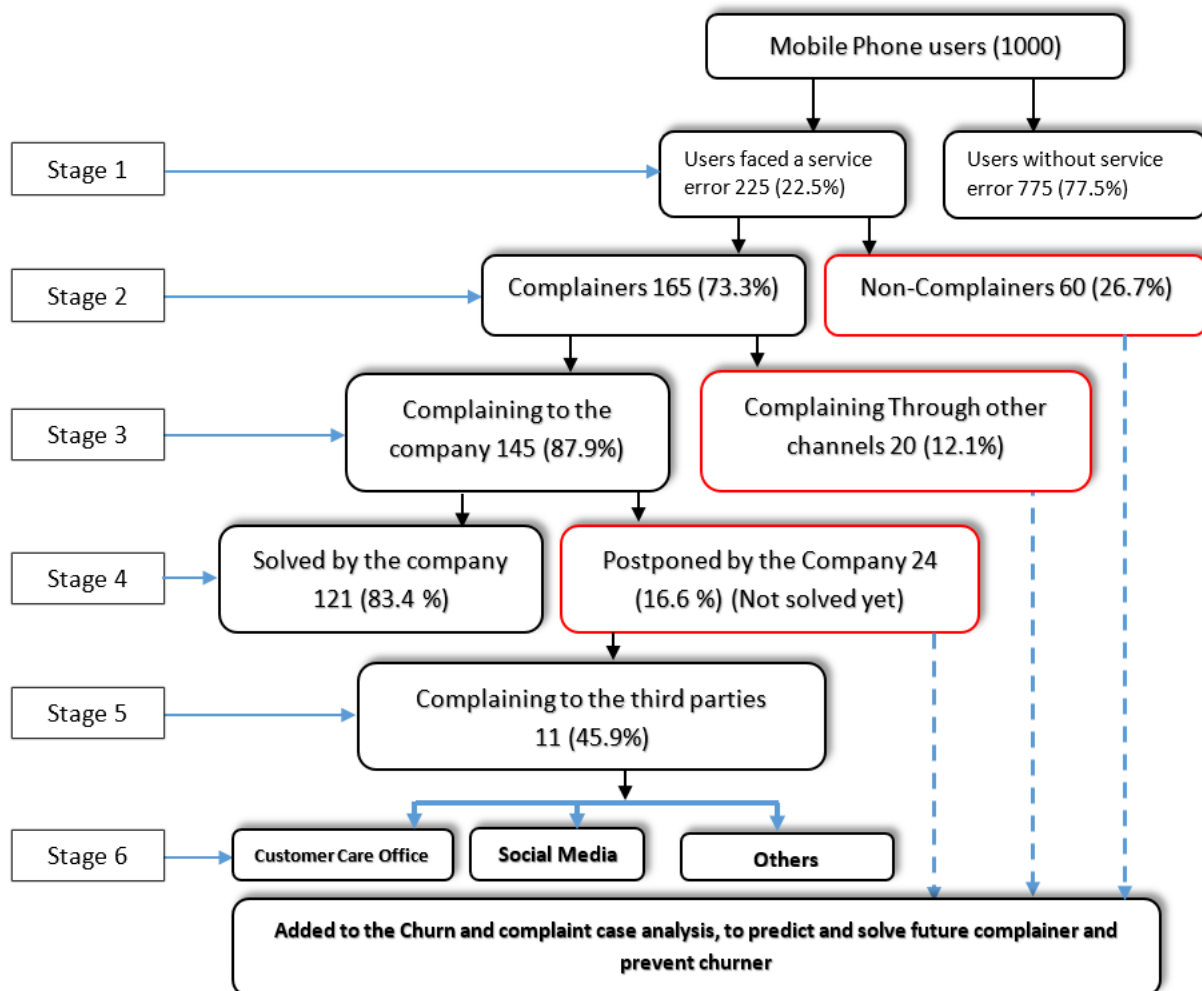
Количественото изследване включва изследване на проблем чрез тестване на теория с помощта на числени данни и статистически анализ. Целта е да се определи дали прогнозите на теорията са общоприложими. Например, изграждането на модел за прогнозиране на отпадането на клиенти в телекомуникационния сектор с помощта на извличане на данни и сложни статистически методи попада в рамките на количественото изследване.

Методологиите на изследване могат да бъдат категоризирани като дедуктивни или индуктивни. Целта на индуктивното изследване е да се развият идеи от конкретни наблюдения, докато дедуктивното изследване е структурирано така, че да тества хипотези, основани на предварително съществуващи убеждения. В това изследване се използват както индуктивни, така и дедуктивни методологии, за да се анализира и създаде модел за прогнозиране на отпадането на клиенти при доставчиците на телекомуникационни услуги, използвайки

данни от обучението. След това моделът се оценява и подобрява, използвайки тестови данни.

➤ Мерки за отлив на клиенти и оплаквания от клиенти

Използвайки набор от данни от 1000 клиенти, набрани в продължение на шест месеца, резултатите показват, че много от клиентите, използващи услугите през деня, не са напуснали услугата и са продължили да я използват в продължение на шест месеца. Освен това, констатациите показват корелация между оплакванията на клиентите и отпадането на клиенти, където събитията на отпадане често съвпадат с оплаквания, което показва, че клиентите без отпадане като цяло са доволни от услугата. За разлика от това, тези с оплаквания са по-склонни да се откажат.



Фигура 13 илюстрира последователния процес в жизнения цикъл на жалбите на клиентите, като подчертава сложността на управлението на

разрешаването на жалбите и необходимостта от стратегии за интервенция, основани на данни, използвайки набор от данни от 1000 клиенти, набрани в продължение на шест месеца по време на случай на неуспех при неуспешно изпращане на SMS с кратък код за абонаменти за услуги (интернет пакет, презареждане, кредитен превод). Резултатите показват, че много клиенти, използващи услугите, не се оплакват директно от компанията. Чрез просто изчисление се заключава, че само приблизително 64,43% от клиентите, изпитващи проблеми с услугата, предоставят отрицателна обратна връзка на фирмата, което показва, че 35,57% от тези клиенти нямат регистрирани оплаквания в базата данни на компанията. Този резултат се получава чрез умножаване на процента на клиентите, които се оплакват от услуга (73,3%), с процента на тези, които са съобщили директно оплакванията си на компанията.

Процент клиенти, които дават отрицателна обратна връзка на компанията (%)

$$= 0.733 \times 0.879 = 0.644 = 64.43\%$$

Уравнение 2. Процент клиенти, изпитващи проблеми с обслужването

Чрез изваждане на този процент от 100%, представляващ общия процент клиенти, изправени пред проблеми с обслужването, може да се покаже, че приблизително 35,57% от клиентите избират да не се оплакват директно от компанията, когато срещнат проблем с обслужването. Вместо това, те могат тихомълком да преминат към друга компания и потенциално да предупредят други клиенти, разпространявайки негативни настроения относно продукта и телекомуникационната компания.

Това въздействие може да стане очевидно за компанията едва по-късно чрез статистически данни. Освен това изчисляването на процента на клиентите, които се оплакват директно на компанията, показва, че само 14.5 % от клиентите избират да съобщават оплакванията си по този начин. Това изчисление се получава чрез умножаване на процента на клиентите, срещащи проблеми с обслужването, с процента на клиентите, които се оплакват под каквато и да е форма, и след това с процента на клиентите, които се обаждат конкретно на компанията, за да се оплачат.

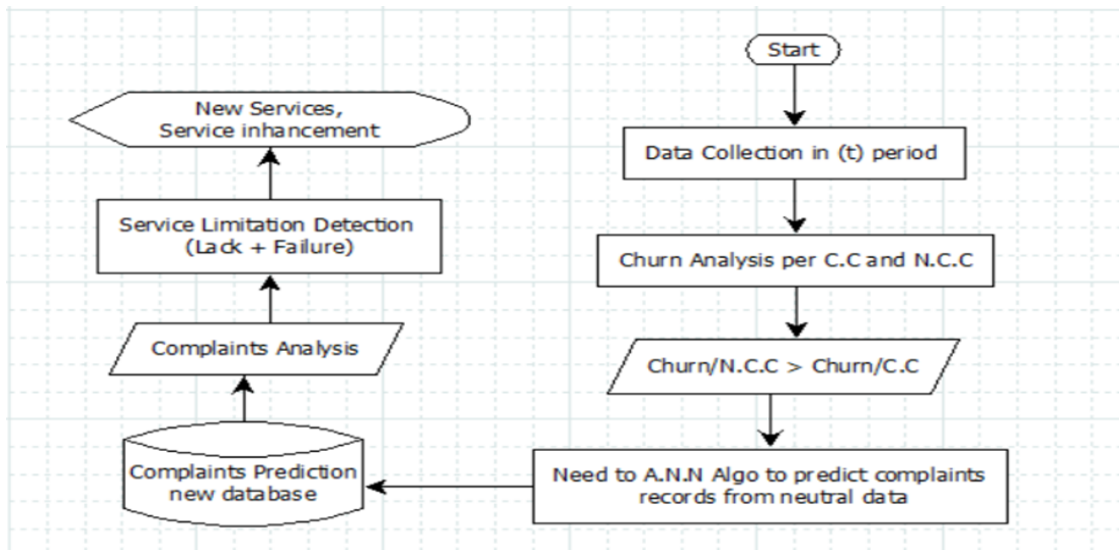
Процент клиенти, които срещат проблеми с обслужването
и подават оплаквания директно до компанията(%)

$$= 0.225 \times 0.733 \times 0.879 = 0.145 = 14.5 \%$$

Уравнение 3: Процент клиенти, които срещат проблеми с обслужването и докладват на компанията

Оставащият въпрос обаче е какво да кажем за останалите 85,5% от клиентите? Доволни ли са от услугата? Или обмислят смяна на доставчика?

Въз основа на гореспоменатото проучване, анализът на оплакванията на клиентите трябва да се фокусира върху идентифициране на причините за ограничения и неуспехи в услугите, като например проблеми с услугите, устройствата и покритието, за да се справят ефективно с тях.



Фиг. 14. Показване на стъпките на анализа на отпадането на клиенти за прогнозиране на оплакванията на клиентите

Фигура 14 илюстрира аналитичните стъпки, предприети систематично за идентифициране на модели на оплаквания, което осигурява методологичната основа за прогнозно моделиране в следващите глави.

Следователно е важно да се обясни, че в сценарии, при които отливът на клиенти със записи за оплаквания (CH/CC) надвишава отлива без оплаквания (CH/NCC), което показва, че $CH/NCC > CH/CC$, става наложително да се разработи нова система за обучение, насочена към прогнозиране на записите за оплаквания на клиенти в цялата база данни.

Въпреки че бяха поучителни, по-ранните проучвания често разчитаха на данни от анкети и интервюта, вместо да събират всички данни за клиентите в единен изчерпателен модел.

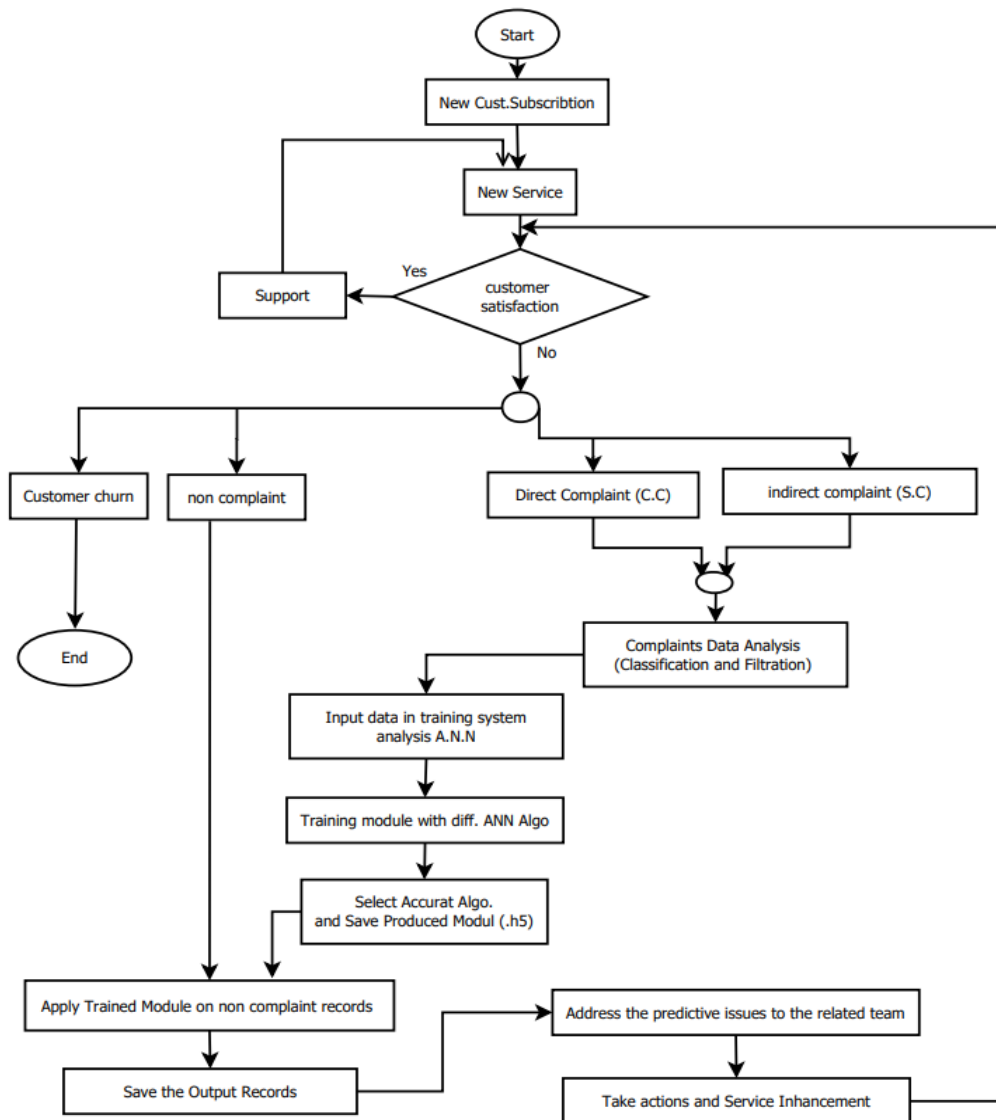
➤ Изследователска стратегия и диаграма

Това проучване ще създаде прогнозен модел, използващ алгоритми за машинно обучение с надзор. (напр. Logistic Regression, Random Forest и модели на невронни мрежи), за прогнозиране на оплакванията на клиентите в сектора на мобилните телекомуникации, а именно на ливанския пазар.

Диаграмата по-долу илюстрира цялостния процес на откриване, анализ и прогнозиране на оплаквания от клиенти в мобилна телекомуникационна среда, използвайки контролирано машинно обучение (ML), а именно изкуствени невронни мрежи (ANN).

Основната цел на тази блок-схема е да илюстрира архитектура за прогнозен анализ, която:

- Идентифицира потенциални предстоящи оплаквания,
- Използва контролирано машинно обучение (ANN) за прогнозиране,
- Повишава удовлетвореността на клиентите чрез предварително разрешаване на проблемите,
- Намалява отлива на клиенти от телекомуникационните услуги.



2.3 Процес на откриване на знания в бази данни

Данните за клиентите на мобилни услуги включват лични данни и показатели за качество на услугата, съхранявани в множество бази данни. Чувствителната информация е криптирана за поверителност. Записите за оплаквания се регистрират в бази данни на колцентрове, докато данните за услугите се събират от мрежови системи. Интегрирането на тези набори от данни подпомага прогнозирането на оплакванията, но остава сложен. Крайният набор от данни е балансиран (50% оплаквания, 50% без оплаквания) и разделен на набори за обучение (60%), валидиране (20%) и тестване (20%) за разработване на модел.

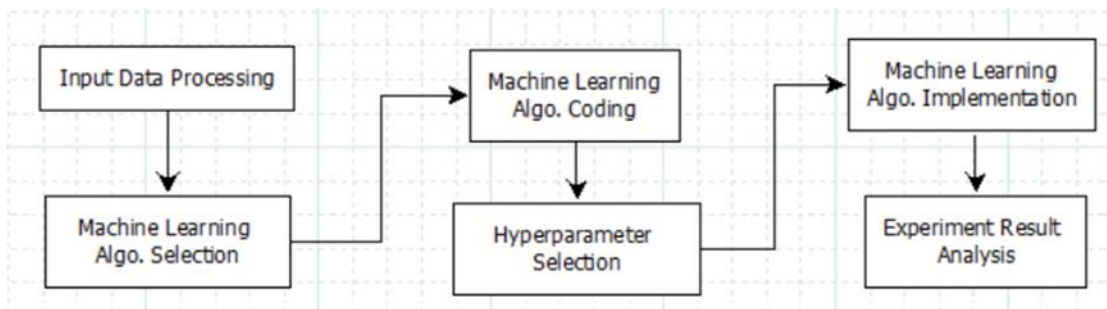
2.4 Алгоритъм за машинно обучение

След провеждане на обширен преглед на литературата, ние идентифицирахме различни техники за прогнозиране на оплакванията на клиентите чрез прогнозиране на отпадането на клиенти в индустрията на мобилните телекомуникации. В резултат на това избрахме четири алгоритъма за това изследване: Logistic Regression, Random Forest и Модели на невронни мрежи (ANN). Тези алгоритми ще бъдат използвани в експерименталната фаза, за да се оцени тяхната точност при прогнозиране на оплакванията на клиентите.

2.5. Дефиниция на модела

Представена е рамката на експеримента, използвана в проучването. Експериментът започва с оценка на структурата и конфигурацията на суровите данни, за да се получи цялостно разбиране за набора от данни.

Тази начална фаза включва и предварителна обработка на данните, за да се гарантира тяхната пригодност за експериментите. Разглеждат се различни алгоритми за машинно обучение въз основа на разкрития от прегледа на литературата. Предприема се задълбочено проучване на различни алгоритми за машинно обучение, за да се направи изборът. След като алгоритмите бъдат избрани, те се имплементират с помощта на Python и библиотеката Scikit-learn. Допълнителни библиотеки в Python, като NumPy, SciPy и Matplotlib, предлагат обширни функционалности за задачи за класификация и регресия. TensorFlow се използва и за реализация на изкуствени невронни мрежи, благодарение на своите обширни пакети за невронни мрежи. След избора и реализацията на алгоритмите за машинно обучение, следващата стъпка включва избиране на хиперпараметри за всеки алгоритъм. В дисертацията се провежда машинно обучение с учител, като всички набори от данни са обозначени с входни променливи (x) и изходна променлива (Y). Алгоритмите се използват за изучаване на функцията за картографиране от входа към изхода, която се обозначава като $Y = f(x)$.



2.6. Заключение

Тази глава описва методологията на изследване, започвайки със събиране на данни и операции по предварителна обработка, последвани от избор и конфигуриране на модел за машинно обучение. Използвайки структурирани канали за данни и възпроизводими работни потоци за обучение, методологията гарантира, че изходът на модела е възпроизводим и валиден. Фигурите предоставят подробна илюстрация на жизнения цикъл на машинното обучение, приложен към прогнозиране на оплакванията на клиентите. Следващата глава прилага тази методология към реален проблем.

Глава 3 Изпълнение

3.1 Архитектура на приложението

Невронна мрежа (ANN), което е съкращение от изкуствена невронна мрежа (ARM), е модел, използван за обработка на информация. Тя е съставена от взаимосвързани неврони, вдъхновени от биологични невронни мрежи. Всяка връзка между изкуствените неврони действа като опростена версия на синапс и може да предава сигнали. Когато бъде получен сигнал, изкуствен неврон го обработва и след това изпраща сигнали към други свързани изкуствени неврони. ANN има множество приложения за проблеми с прогнозирането и класификацията в индустриалните и бизнес области.

Ако искаме да обучим нашия модел да отчита теглата и отклоненията, процедурата за обучение работи по следния начин:

1. Инициализираме теглата за всички възли.
2. Извършваме предварително преминаване, използвайки текущите тегла за всеки пример за обучение, изчислявайки изхода на всеки възел (стойността на последния възел).
3. След това сравняваме крайния резултат с действителния етикет в данните за обучение (обучение с учител) и измерваме грешката, използвайки функция на загуба.

4. След това извършваме обратно преминаване от дясно на ляво, разпространявайки грешката до всеки възел, използвайки обратно разпространение.
5. Изчисляваме приноса на всяко тегло към грешката и коригираме теглата и отклоненията съответно, използвайки градиентен спускащ метод.

3.2 Поток от данни и обработка

Суровите данни са склонни към шум, липсващи стойности и несъответствия. Качеството на данните значително влияе върху резултатите от машинното обучение. Предварителната обработка на данните е от решаващо значение за подобряване на качеството на данните и повишаване на ефективността на процеса на извличане на данни. Методите за предварителна обработка на данни включват почистване на данни, интегриране на данни, трансформация на данни и редукция на данни.

Уравнение 9 представлява опростената формула за min-max нормализация, където $\max d_j$ е максималната стойност на променливата, а $\min d_j$ е минималната стойност на променливата.

$$d'_i = \frac{d_i - \min d_j}{\max d_j - \min d_j}$$

Уравнение 9: Опростена формула за минимално-максимална нормализация

3.3 Описание на набора от данни

Тази база данни беше използвана за събиране на необходимата информация за по-нататъшен анализ. По-долу е представен примерен набор от данни.

Описания на входните колони:

- Модел на устройството: При осъществяване на гласови или информационни повиквания в локална или международна мрежа
- Област: Когато клиент подаде запис на обаждане,
- PSTN_ID: Обществена комутирана телефонна мрежа (PSTN)
- Възраст: Възраст на клиента
- Заглавие: Използва се като типичен маркер за големи клиенти
- Локално повикване: Когато повикването е в рамките на локалната мрежа
- Роуминг: Когато разговорът е в международната мрежа

- СРЕДНА продължителност/цикъл на гласови повиквания: Средната продължителност на локалните гласови повиквания в рамките на един и същ цикъл на фактуриране на клиента
- СРЕДНА продължителност/цикъл на разговори за данни: Средната продължителност на локалните разговори за данни се измерва в рамките на едно и също клиентско фактуриране.
- Брой гласови повиквания/цикъл: Броят на локалните гласови повиквания в рамките на един и същ цикъл на фактуриране на клиента
- Брой повиквания за данни/цикъл: Брой локални повиквания за данни в рамките на един и същ цикъл на фактуриране на клиента
- Тип услуга: Типът услуга се базира на скоростта на предаване на данни, 3G или 4G LTE.
- Скорост: Отнася се до скоростите на предаване на данни, като например 3G или 4G LTE.
- Проблем с трафика на мрежовото покритие: Проблемът е регистриран, когато гласово повикване за конкретен клиент е неуспешно.
- Грешка при консумация на данни: Проблемът е регистриран, когато конкретен клиент не успява да се свърже с услуга за пакет данни със скорости 3G или 4G LTE.

3.4 Подготовка за оплаквания и набори от данни за отлив на клиенти

Целта на този модел е да идентифицира основни атрибути на предсказващите фактори, които помагат на бизнеса да разбере индикаторите за отпадане на клиенти. Логистичен регресионен модел или модел от семейството на дърветата е отличен избор за идентифициране на значими променливи.

Данните са събирани в продължение на 6 месеца, като е оценена информацията относно клиентите, както е показано в следващата таблица. Този набор от данни е използван допълнително за обучение на модели за машинно обучение.

Таблица 11. Пример за набор от данни за отпадане

Зона	Подномер	Код на областта	Между народ ен план	План за VMail	Съобщени е от VMail	Дневни минути	Дневни разговор и	Дневна такса
зона017	157832276	415	не	да	25	265.1	110	45.07
зона034	142891510	415	не	да	26	161.6	123	27.47
зона030	7580360	415	не	не	0	243.4	114	41.38

зона034	142750572	408	да	не	0	299.4	71	50.9
зона035	7429276	415	да	не	0	166.7	113	28.34
зона002	153871838	510	да	не	0	223.4	98	37,98
зона020	7670850	510	не	да	24	218.2	88	37.09
зона024	142617110	415	да	не	0	157	79	26.69

→

Ив Мин с	Ив зове	Заряд на Ив	Нощ ни мину ти	Нощн и обажд ания	Нощн о зарез дане	Ме жду нар одн и мин ути	Меж дуна родн и разго вори	Межд унаро дна такса	Обаждани я за обслужван е на клиенти	Отлив	оплакван е
197.4	99	16.78	244.7	91	11.01	10	3	2.7	1	Фалши во.	0
195.5	103	16.62	254.4	103	11.45	13.7	3	3.7	1	Фалши во.	1
121.2	110	10.3	162.6	104	7.32	12.2	5	3.29	0	Фалши во.	1
61.9	88	5.26	196.9	89	8.86	6.6	7	1.78	2	Фалши во.	1
148.3	122	12.61	186.9	121	8.41	10.1	3	2.73	3	Фалши во.	0
220.6	101	18.75	203.9	118	9.18	6.3	6	1.7	0	Фалши во.	0
348.5	108	29.62	212.6	118	9.57	7.5	7	2.03	3	Фалши во.	0
103.1	94	8.76	211.8	96	9.53	7.1	6	1.92	0	Фалши во.	0
351.6	80	29,89	215.8	90	9.71	8.7	4	2.35	1	Фалши во.	0

Процесът трансформира категоричните данни, така че те да могат да се използват ефективно при обучение на модели от невронна мрежа, както и да бъдат съвместими със слоевете Keras. Поддържа се таблица за картографиране на кодирани изходи, за да се преобразуват кодираните изходи в действителни изходи по време на моделиране.

3.5 Внедряване на машинно обучение

Прикачен Python код LogisticReg.py

```
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.utils.class_weight import compute_class_weight
from sklearn.metrics import (
    accuracy_score, precision_score, recall_score,
    f1_score, confusion_matrix, classification_report,
    precision_recall_curve, average_precision_score
)
from imblearn.over_sampling import SMOTE
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np

# Load dataset
dataset = np.loadtxt('dataSetLast.data.csv', delimiter=',')
x = dataset[:,1:15]
y = dataset[:,15].ravel() # Flatten y to shape (n_samples,)

# Train/test split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.20, random_state=42)

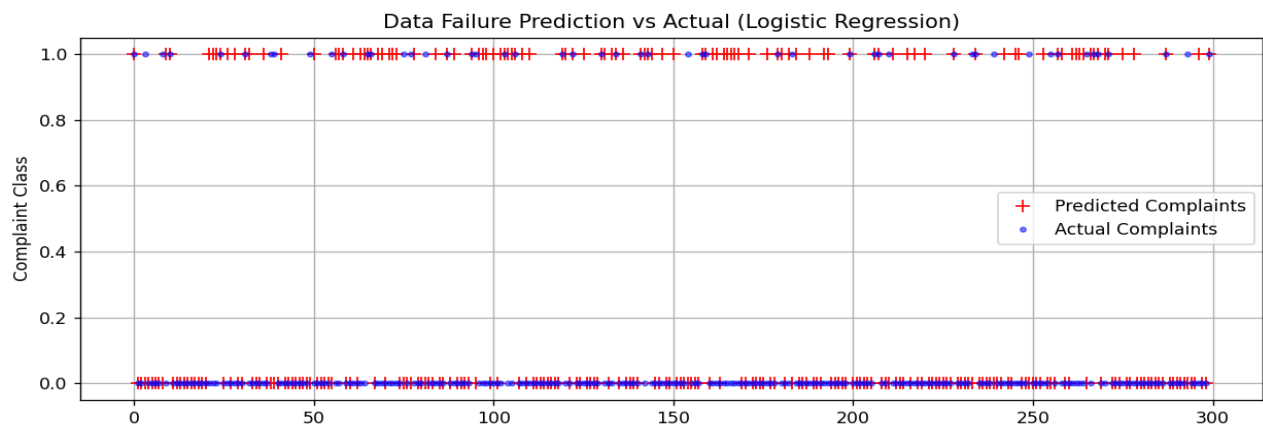
# Apply SMOTE
smote = SMOTE(random_state=42)
x_train_sm, y_train_sm = smote.fit_resample(x_train, y_train)

# Train Logistic Regression
model = LogisticRegression(max_iter=1000)
model.fit(x_train_sm, y_train_sm)

# Predict
y_pred = model.predict(x_test)
y_scores = model.predict_proba(x_test)[:, 1]

# Metrics
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred)
recall = recall_score(y_test, y_pred)
f1 = f1_score(y_test, y_pred)

print(f"Accuracy: {accuracy:.2f}")
print(f"Precision: {precision:.2f}")
print(f"Recall: {recall:.2f}")
print(f"F1 Score: {f1:.2f}")
print("\nClassification Report:\n", classification_report(y_test, y_pred, target_names=['No Complaint', 'Complaint']))
```

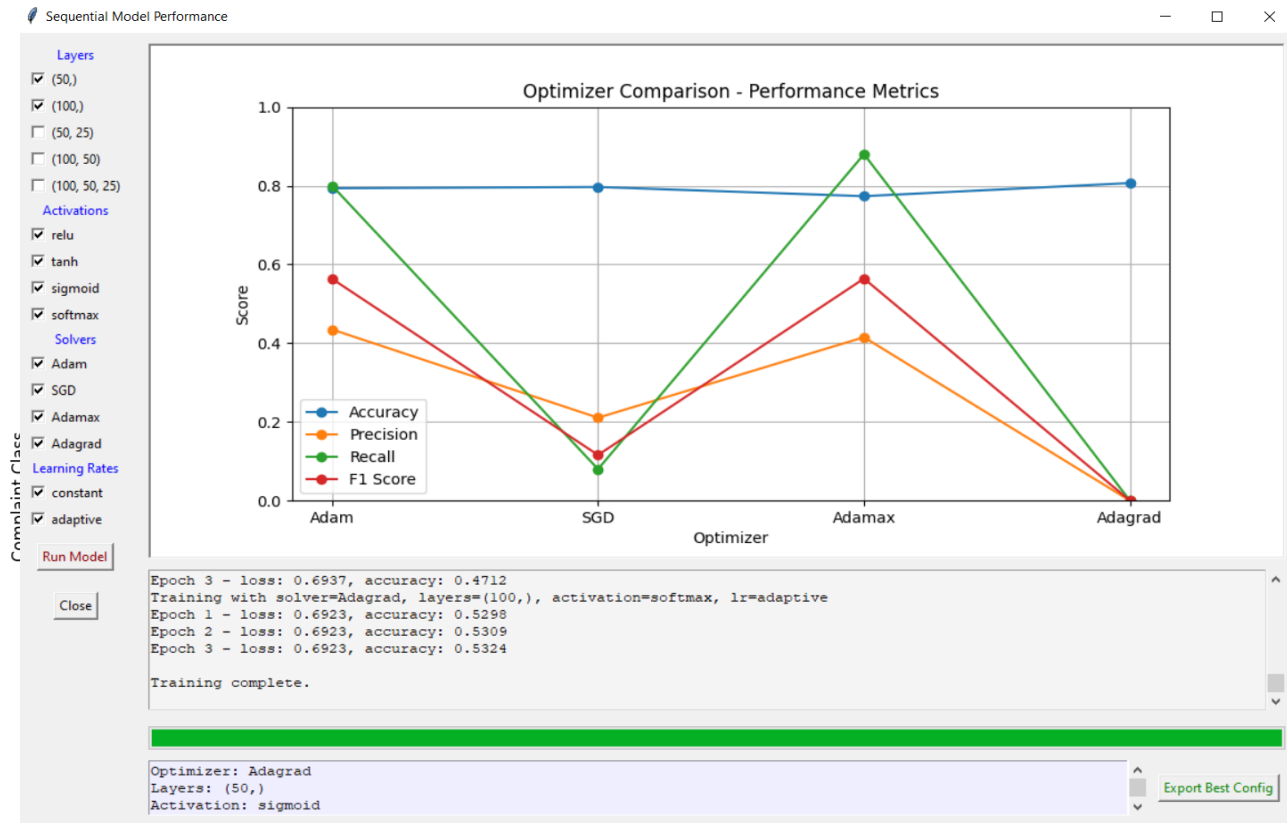


Прикачен код на RandomForest.py:

```
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn.utils.class_weight import compute_class_weight
from sklearn.metrics import (
    accuracy_score, precision_score, recall_score,
    f1_score, confusion_matrix, classification_report,
    precision_recall_curve, average_precision_score
)
from imblearn.over_sampling import SMOTE
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np

# Load dataset
dataset = np.loadtxt('dataSetLast.data.csv', delimiter=',')
x = dataset[:,1500, 1:15]
y = dataset[:,1500, 16:17].ravel() # Convert y to 1D

# Train/test split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.20, random_state=42)
```

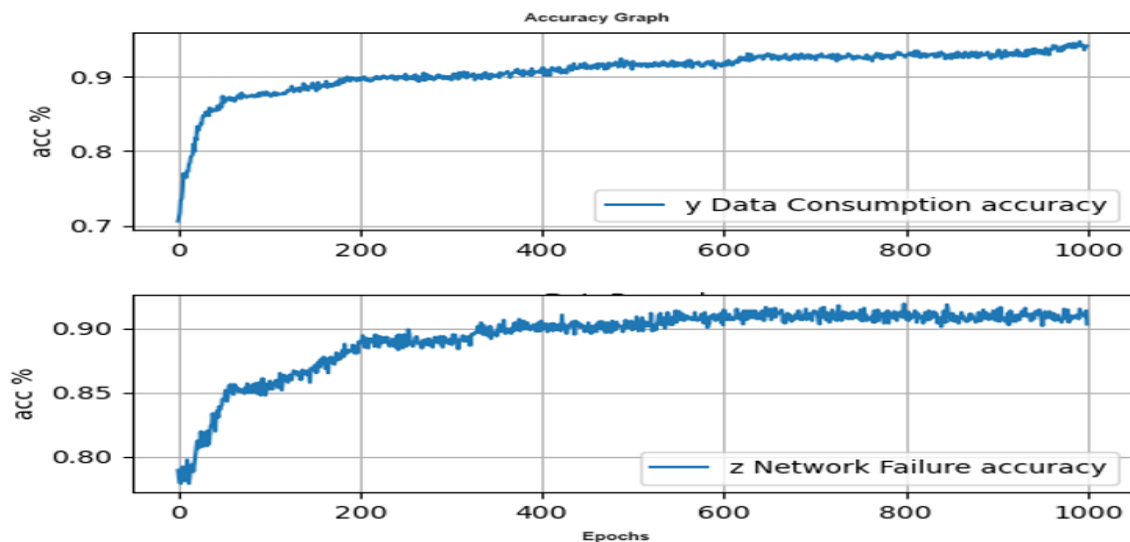


В интерфейса по-долу представяме сравнителна оценка на четири различни оптимизатора, **Adam**, **SGD**, **Adamax** и **Adagrad**, въз основа на тяхната производителност в рамките на модел на изкуствена невронна мрежа (ANN) за прогнозиране на оплаквания в телекомуникационните компании. Оптимизаторите се оценяват с помощта на четири стандартни показателя за класификация: **Точност**, **Прецизност**, **Извикване** и **F1 резултат**.

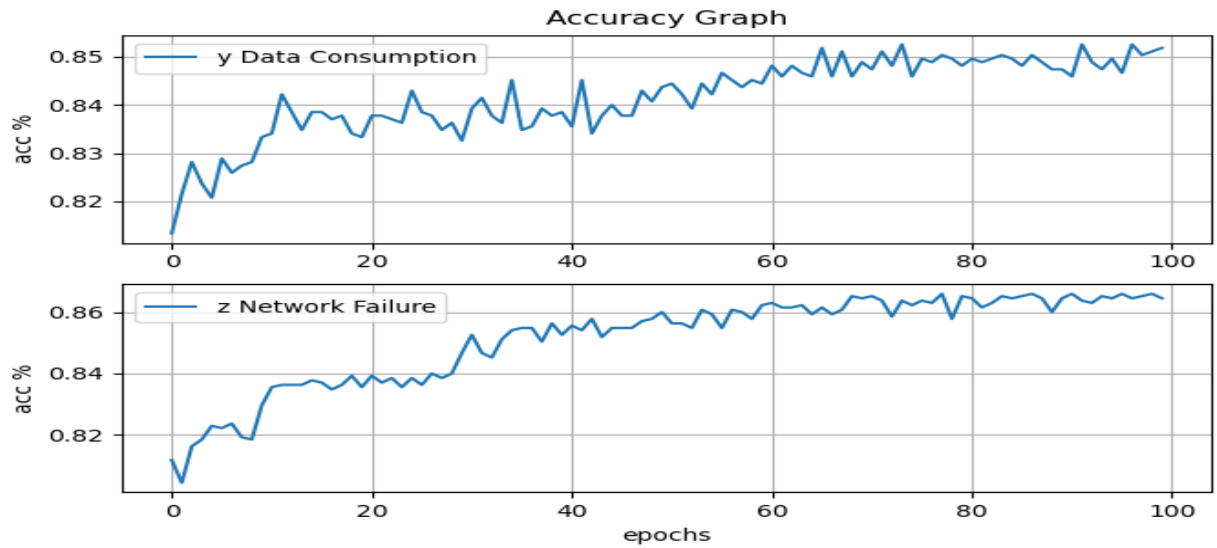
Показаният интерфейс предоставя графична среда за оценка на производителността на различни оптимизатори в модел на последователна изкуствена невронна мрежа (ANN). Той позволява на потребителите да

конфигурират хиперпараметри на модела, като например структури на слоевете, активиращи функции, алгоритми за оптимизация и скорости на обучение. След като моделът е обучен, системата представя сравнителна диаграма на производителността, базирана на ключови показатели за класификация: точност, прецизност, извиканост и F1 резултат. Интерфейсът показва също така дневници за обучение в реално време, проследяване на напредъка и подчертава най-добре представящата се конфигурация. Този инструмент улеснява емпиричното настройване на модела и подкрепя информираното вземане на решения при оптимизация на невронни мрежи за задачи като прогнозиране на оплаквания от клиенти.

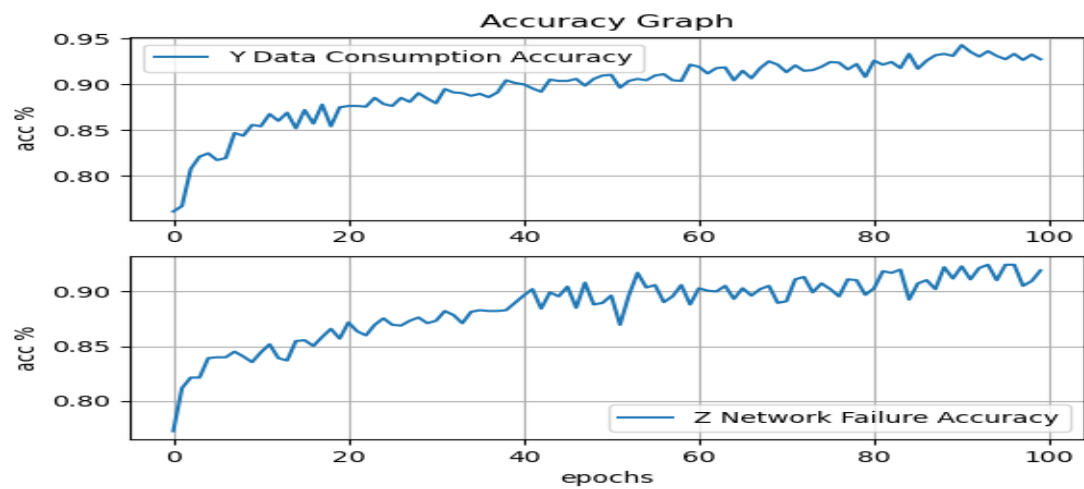
Скорост на обучение на оптимизатора ADADELTA = 0.1, Епохи = 1000



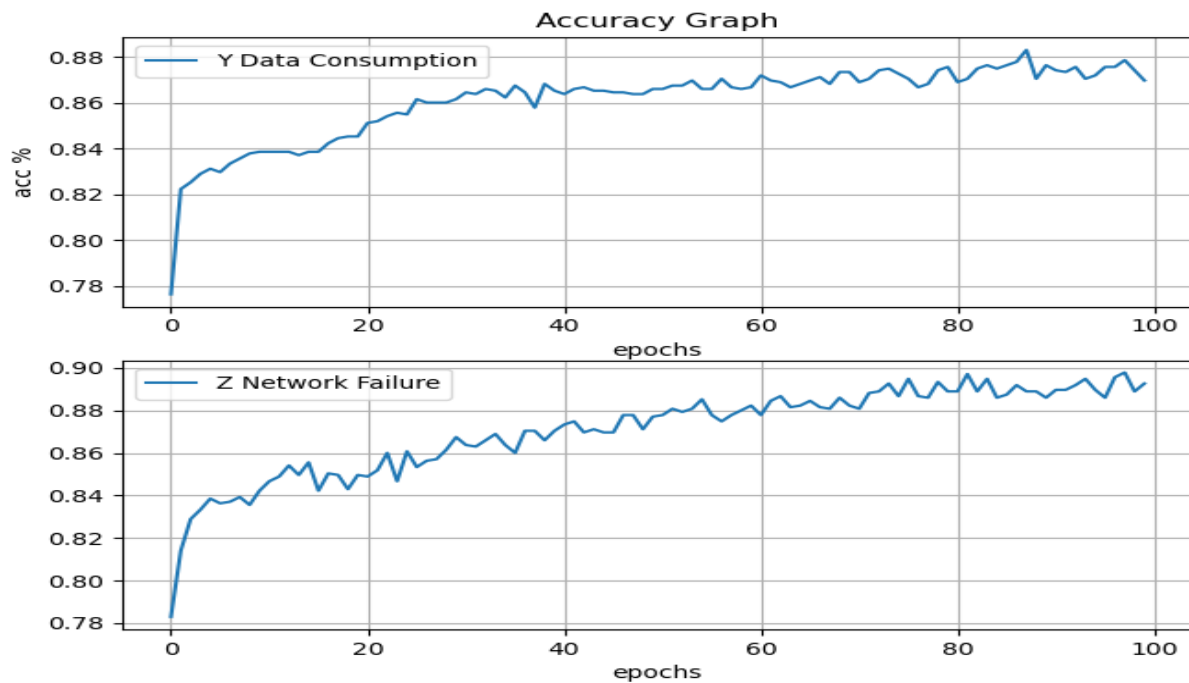
Скорост на обучение на оптимизатора ADAGRAD = 0.01, Епохи = 100



Скорост на обучение на ADAMAX Optimizer = 0.01, Епохи = 100

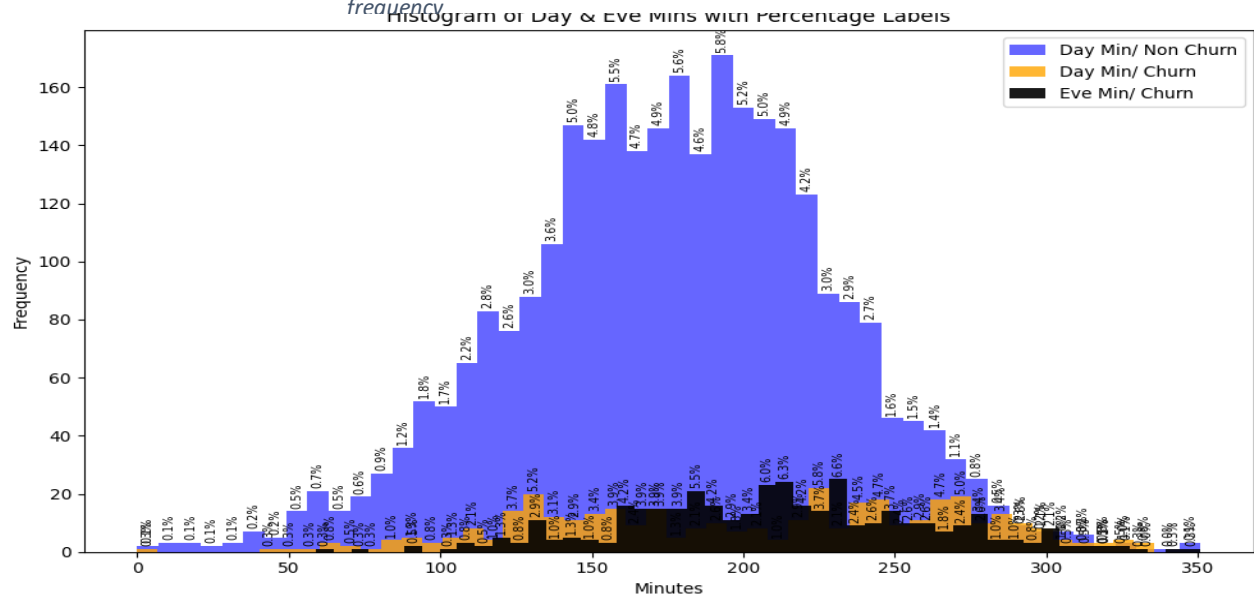


Оптимизатор на скоростта на обучение на RMSprop = 0.001, Епохи = 100



Случай на употреба в телекомуникационния сектор
(анализ на отпадане на клиенти спрямо оплаквания)

Figure SEQ Figure * ARABIC 48 Customer activities for day call frequency



Тази хистограма визуализира разпределението на минутите за разговори между различните клиентски сегменти, като по-специално анализира дневните минути и вечерните минути на клиентите, които са напуснали и не са напуснали.

Тази хистограма показва, че **моделите на продължителност на обажданията се различават между клиентите, които са напуснали компанията, и тези, които са запазили компанията.** Ненапускащите клиенти са склонни да показват по-последователно и умерено поведение при обаждания през деня, докато напускащите клиенти имат по-разпръснат и разнообразен модел на употреба, както през деня, така и вечерта. Тези тенденции в употребата биха могли да служат като предсказващи характеристики в моделите за класификация на напускащите клиенти.

3.6 Заключение

Главата за внедряването трансформира методологичните и теоретичните основи в оперативен модел. Тя описва как изкуствените невронни мрежи са създадени, проектирани и обучени да предсказват оплакванията. Различни конфигурации на модела, включително оптимизатори и слоеве за отпадане, са експериментирани, за да се постигне по-добра производителност. Диаграмите помогнаха за описанието на архитектурните избори и стратегиите за обработка на данни. Последиците и резултатите от тези внедрявания са разгледани в следващата глава.

Глава 4 Резултати и оценка

Това проучване използва реални телекомуникационни данни, за да предскаже оплакванията на клиентите чрез машинно обучение, постигайки точност над 92%. Най-добрият модел използва два плътни слоя с ReLU, един със Softmax и оптимизатора ADADELTA с коефициент на обучение 0,1. Изкуствените невронни мрежи превъзхождат други методи и допълнителни ползи могат да дойдат от по-големи набори от данни и настройване на комбинациите от скорост на активиране и обучение.

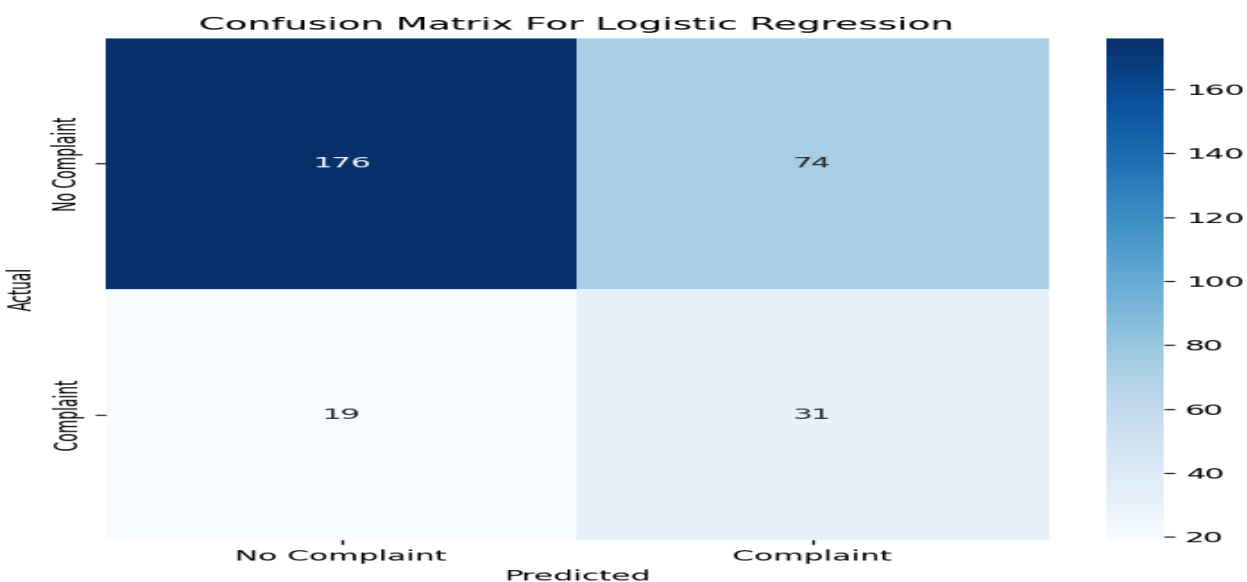
4.1 Анализ на модели на машинно обучение Logistic Regression, Random Forest и ANN

- Logistic Regression (LR) и анализ на показатели

Предоставената матрица на смесване визуализира ефективността на модела Logistic Regression при прогнозиране на оплаквания от клиенти на телекомуникационни компании. Матрицата сравнява прогнозираните от модела резултати с действителните резултати за клиентите.

Матрицата съдържа четири квадранта:

1. Истински положителни резултати (31 случая):
Правилно е било предвидено, че клиентите ще имат оплаквания.
2. Фалшиво положителни резултати (74 случая):
Клиентите неправилно са предвидили, че имат оплаквания (всъщност няма оплакване).
3. Истински отрицателни резултати (176 случая):
Клиентите правилно са предвидили, че нямат оплаквания.
4. Фалшиво отрицателни резултати (19 случая):
Клиентите неправилно са предвидили, че нямат оплаквания (всъщност са имали оплаквания).



Фигура 1: Графика с резултати на модела Logistic Regression

➤ Модел Random Forest и анализ на резултатите

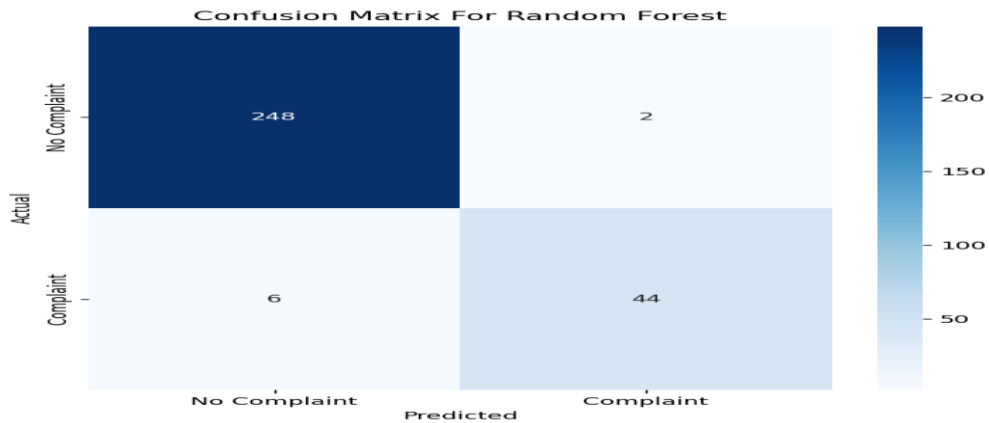


Figure 53 Random Forest Confusion Matrix

Този матрица на смесване обобщава колко ефективно моделът Random Forest е предсказал оплакванията на клиентите в телекомуникационен сценарий:

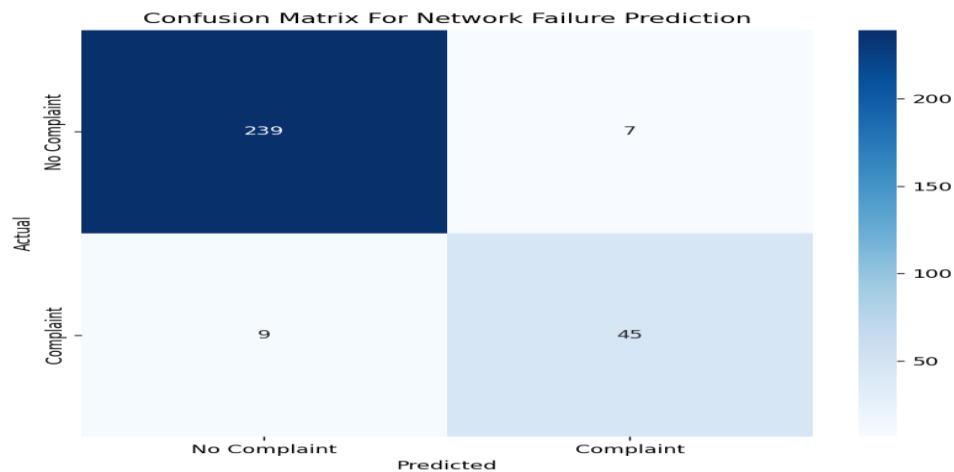
- 248 Истински отрицателни: Правилно предсказани клиенти, които нямат оплаквания.
- 44 истински позитива: Правилно прогнозирани клиенти, които са имали оплаквания.
- 2 фалшиви положителни резултата: Неправилно прогнозиране, че клиентите ще имат оплаквания, когато те не са имали такива.
- 6 фалшиви отрицателни резултата: Неуспешно предвиждане на клиенти, които действително са имали оплаквания.

➤ Модел на невронна мрежа (ANN)

Фигура 57 илюстрира производителността на модела за прогнозиране на оплаквания за мрежови повреди. Моделът постигна висока точност на класификация, като правилно предсказа:

- **239 истински отрицателни резултата** (клиенти без оплаквания)
- **45 истински положителни резултата** (клиенти с реални оплаквания)
- **7 фалшиво положителни резултата** (клиенти, които не се оплакват, се очаква да се оплакват)
- **9 фалшиво отрицателни резултата** (не са открити действителни оплаквания)

Тези резултати показват, че моделът демонстрира силна предсказваща способност, особено при идентифициране на случаи без оплаквания (истински отрицателни), като същевременно поддържа разумен баланс при откриване на действителни оплаквания (истински положителни). Това подкрепя неговата полезност при управлението на оплакванията и оптимизирането на обслужването на клиентите в областта на телекомуникациите.



Фигура 27 Матрица на смесването за точността на прогнозиране на жалбите

4.2 Статистически анализ на случаите на употреба в телекомуникациите за модели на продължителност на разговорите

Анализът на продължителността на дневните разговори за период от шест месеца показва, че клиентите, които не са се отказали от услугата, са имали по-висока честота на обажданията и по-дълга продължителност, което показва по-голяма ангажираност и удовлетвореност. За разлика от тях, клиентите, които са се отказали от услугата, са показали по-ниска активност, което предполага недоволство или по-добри алтернативи. Тези модели могат да помогнат на телекомуникационните компании да идентифицират клиенти в риск и да разработят целенасочени оферти или кампании за лоялност, за да подобрят задържането на клиенти и удовлетвореността от обслужването.

4.3 Статистически анализ на модела за отлив на клиенти и оплаквания

Наборът от данни проследява поведението на клиентите в продължение на шест месеца и идентифицира три фази на отпадане: Добра, Действие и Отпадане. Повечето клиенти показват недоволство във фазата на действие, преди да напуснат. Анализът разкрива силна корелация между оплакванията и отпадането - много от тези, които напускат, са се оплаквали и преди. Около 15% обаче се оплакват, но не напускат, което предполага, че оплакванията не винаги водят до загуба. Тези прозрения подчертават стойността на проактивното обработване на оплаквания и прогнозния анализ за намаляване на отпадането и подобряване на задържането на клиенти.

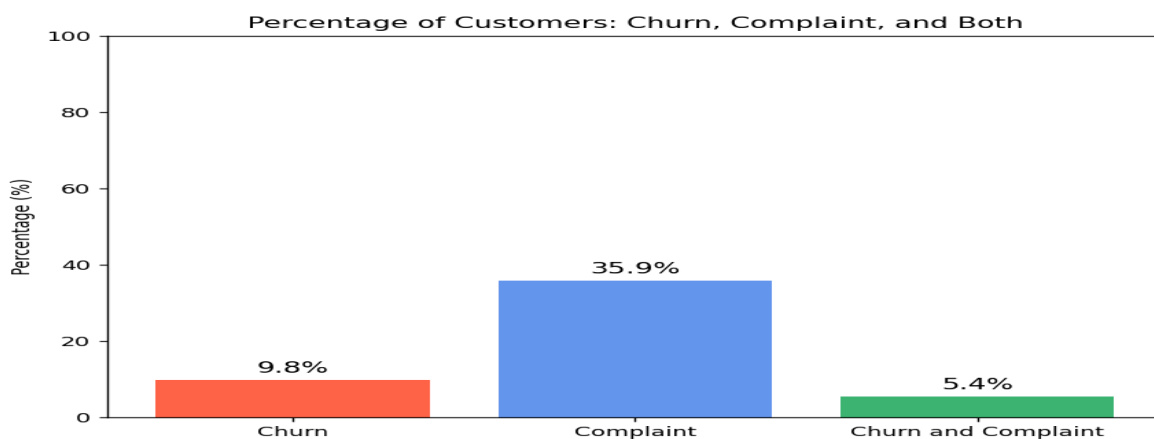


Фигура 62 представя разпределението на поведението на клиентите по отношение на отпадане и оплаквания въз основа на анализа на набора от данни.

- Отпадане (9,8%):
Това показва, че 9,8% от всички клиенти са прекратили услугата си. Макар че това е сравнително малък дял, той отразява общия процент на отпадане и сигнализира за област с потенциална нужда от намеса.
- Жалба (35,9%):
Повече от половината клиенти (35,9%) са подали поне едно оплакване. Това подчертава значително ниво на неудовлетвореност или проблеми с обслужването, които изискват внимание от страна на екипите за обслужване на клиенти и оперативните екипи.
- Отлив и оплаквания (5,4%):

Тази подгрупа представлява 5,4% от клиентите, които са подали оплакване и по-късно са го напуснали. Тези клиенти са особено критични, защото оплакванията им не са били разрешени достатъчно ефективно, за да ги задържат. Те служат като ключов индикатор за пропуснати възможности за задържане на клиенти.

- По-голямата част от потребителите, които не са се включили в поръчката, изглежда си тръгват, без да подават оплакване, което показва **мълчаливо недоволство** или **безпроблемно превключване между клиенти**.



4.4 Резултати и сравнение

Използвахме реални набори от данни от мобилни оператори в индустрията, за да обучим нашия модел. „Y Асигурасу“ показва неуспех в потреблението на данни, а „Z Асигурасу“ показва мрежова повреда. Нашето изследване въведе нов подход, който постига точност на прогнозиране над 92%. Въпреки че тази точност е впечатляваща, най-добрите параметри на модела включват две нива на Dense с активиране „RELU“ и едно Dense с активиране „Softmax“, използвайки оптимизатора ADADELTA с малка скорост на обучение от 0,1. Вярваме, че използването на по-големи набори от данни за обучение и оптимизирането на комбинацията от функции за активиране и свързаните с тях скорости на обучение биха могли да доведат до допълнителни подобрения.

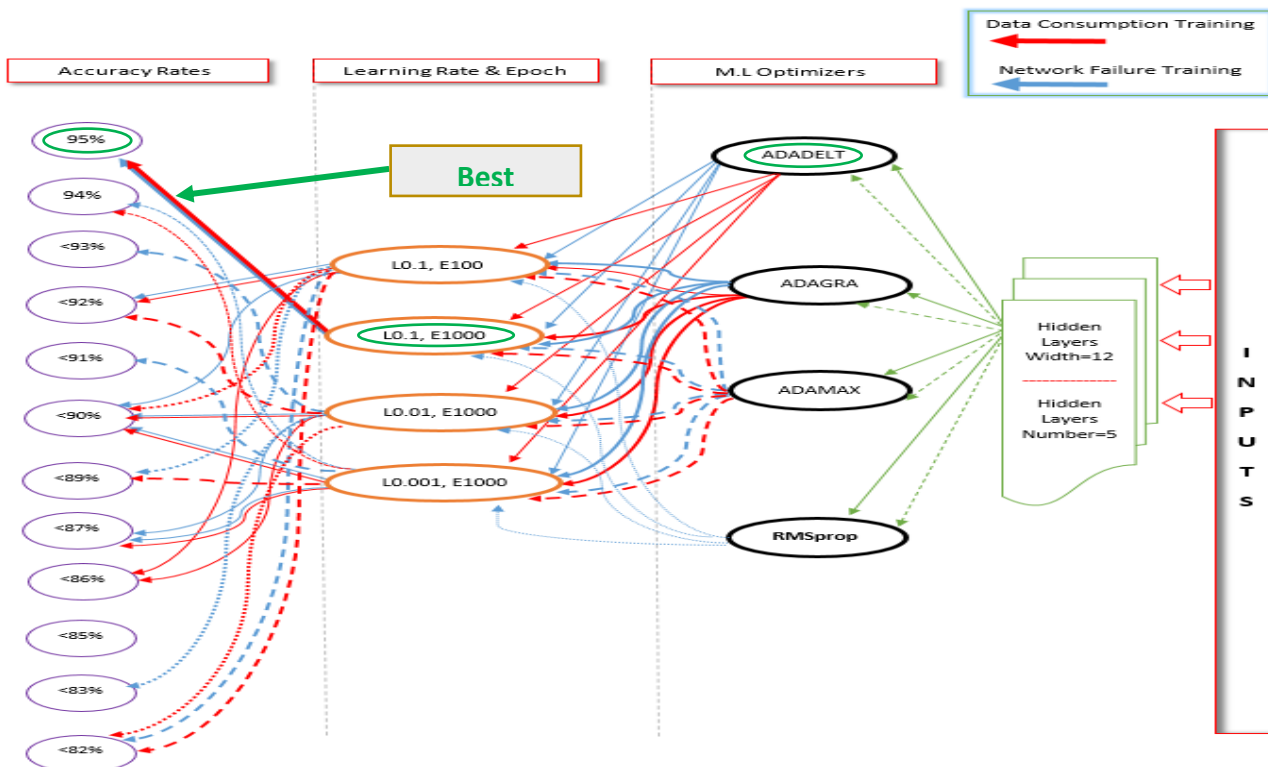


Table 141 Training results with different ML Models

Метричен	Оптимизатор на ANN Adadelata	Logistic Regression	Random Forest
Общо записи	300	300	300
Истински позитиви (ИП)	45	31	44
Истински отрицателни (ТН)	239	176	248
Фалшиво положителни резултати (ФП)	7	74	2
Фалшиво отрицателни резултати (ФО)	9	19	6
Точност	95,67%	69,00%	96,33%
Прецизност	91,54%	29,52%	95,65%
Припомняне	85,33%	62,00%	88,00%
F1-резултат	89,91%	39,95%	91,79%

Таблица 14 показва силните резултати и за двата модела - Random Forest и ANN. При сравнението Random Forest показва точност от 96,33%, а ANN показва точност от 95,67%. Резултатите се считат за сходни, но решението

е да се използва изкуствена невронна мрежа (ANN) заради следните предимства и характеристики:

- Способност за улавяне на сложни нелинейни взаимовръзки: особено тези с множество скрити слоеве са изключително добри в обучението
- Обучение на характеристики (имплицитно): Ако суровите телекомуникационни данни (записи с подробности за обажданията, мрежови лог файлове) имат много добре развити модели, които не са очевидни за човешките инженери, ИНН може да ги открие.
- Мащабируемост с големи набори от данни
- Работа с различни типове данни

Ето защо повечето телекомуникационни компании в крайна сметка биха използвали по-точни и адаптивни решения за прогнозиране на оплакванията на клиентите.

4.5 Резултати и оценка

След изграждане на графики за различните модели на оптимизатор и точни параметри за прогнозиране на оплакванията в Таблица [16] и Фигура 50, този модел за записите на жалбоподателите ще бъде използван за тяхното прогнозиране и вместо да се обработва всеки един поотделно, новата техника за системен анализ и поддръжка ще бъде полезна.

Случай на употреба 1: Прекъсване на трафика при мрежово покритие

Прилагаме системния модел върху избрани клиенти в определена област и регистрираме записаните оплаквания за същите тези клиенти за един ден, за да уловим истински положителните, истински отрицателните и фалшиво положителните, фалшиво отрицателните резултати, за да докажем осъществимостта и точността на този модел за прилагане върху CRM бази данни на Hall.

Таблица 14 Записи на жалби от клиенти по център за обаждания/ден

	Подномер	Идентификатор на устройството	Модел на устройството	Площ	PSTN_ID	Местно обаждаме	Тип услуга	Скорост	Мрежово покритие Трафик Прекъсване	Обяснение на клиента
Преп. №1	76832707	cyIQso	Samsung Galaxy Grand Prime PRO SM-J250FDS	Бейрут	460	ДА	HS3	4G	0	
Преп. №2	71567730	UaEy2U	Samsung SM-G316HUDS /Samsung Galaxy S Duos 3	Бейрут	288	ДА	HS4	4G	1	Слаб сигнал
Преп. №3	76728144	ПАпККИ	Стилус за LG H540/LG G4	Бейрут	545	ДА	HS3	4G	0	
Преп. №4	70667693	qCyr2i	Apple A1905/Apple iPhone 8	Бейрут	396	ДА	XC2	4G	0	
Преп. №5	81945448	b6rzkK	Apple A1981/Apple iPhone 14 Pro Max	Бейрут	546	ДА	HS4	3G	0	
Преп. №6	81791458	tX5OIO	Apple A1577/Apple iPhone 12 Pro	Бейрут	322	ДА	XC2	4G	0	
Преп. №7	3867411	9MU0wl	Samsung SM-G610FDS/ Samsung Galaxy J7 Prime Duos	Бейрут	80	ДА	HS3	4G	0	
Преп. №8	70000540	FJ0sIB	Huawei LND-L29/ Huawei Y6 (2018)	Бейрут	220	ДА	HS4	4G	0	
Преп. №11	76781452	mp6hSG	Huawei VNS-L31/Huawei P9 Lite	Бейрут	20	ДА	HS4	4G	1	Прекъсване на повикванията
Преп. №10	71251499	972liq	Samsung SM-G570FDS/ Samsung Galaxy J5 Prime Duos	Бейрут	548	ДА	XC2	4G	0	

Както наблюдаваме в таблица 17, в регистрираните оплаквания от екипа на колцентъра за един ден, имаме само две оплаквания, свързани с прекъсване на мрежовото покритие, и всяко от тях е представено с различно обяснение от страна на клиента (Рек. №2: Слаб сигнал и Рек. №11: Прекъсване на повикванията).

Таблица 15 Записи на жалби от клиенти по център за обаждания/за един ден

	Call Center Record	Customer Explanation	Prediction Results by our AI-Based Model		
			Predicted Record by our Model	Accuracy	Model Prediction and Technical Explanation
Rec#1	0		0	True Negative	
Rec#2	1	Low signal	1	True Positive	PSTN Issue due the fuel and tank damage
Rec#3	0		0	True Negative	
Rec#4	0		0	True Negative	
Rec#5	0		0	True Negative	
Rec#6	0		1	False Positive	Site issue and trafic redirection
Rec#7	0		0	True Negative	
Rec#8	0		0	True Negative	
Rec#9	0		0	True Negative	
Rec#10	0		0	True Negative	
Rec#11	1	Calls disconnection	1	True Positive	PSTN Issue due the fuel and tank damage
Rec#12	0		0	True Negative	
Rec#13	0		0	True Negative	
Rec#14	0		0	True Negative	
Rec#15	0		0	True Negative	
Rec#16	0		0	True Negative	

В Таблица 18 използваме нашия модел, базиран на изкуствен интелект, върху същите клиентски записи. Наблюдаваме допълнителни резултати и от двете страни: броя на оплакванията и обяснението. В броя има три прогнозирани оплаквания (Рес#2, Рес#11 и Рес#6). Първите две оплаквания са подобни на реалните записи в Таблица 101, но третото (Рес#6) е прогнозирано от нашия модел, базиран на изкуствен интелект, което ни дава знак, че може да имаме много други оплаквания се, че са се сблъскали с повреда в услугата, но не са регистрирани от екипа на кол центъра.

В повечето случаи клиентите не са наясно с коренните причини за възникналия проблем; те могат да обяснят проблема или проблема с услугата с различна

диагностика, което води до забавяне в процеса на намиране на решения и отстраняване на проблема, тъй като причините не са ясни и изискват достатъчно време за разследване. Това създава още един проблем с удовлетвореността на клиентите, подновяването на услугата и финансовите приходи.

Както наблюдаваме, обясненията на клиента в Res#2 и Res#11 в таблица 101 се различават от тези в прогнозната таблица 102, която включва ясни обяснения и е посветена на текущия проблем, което ще помогне за решаването му в кратки срокове и понякога може да се случи, преди да получим оплакване.

Таблица 16 Сравнение на резултатите за точността на модел, базиран на изкуствен интелект, върху данни в реално време

	Т. Положителен	Т. Отрицателен	Ф.Положителен	Ф. Отрицателен
Общо записи = 16	2	13	1	0
	ВЯРНО		НЕВЯРНО	
	15		1	
	Положителен %		Отрицателен %	
	93.75		6.25	

В таблица 103 е показан положителният процент на ТР (истински положителни) и FP (фалшиво положителни) е 93,75%, което се счита за добър процент, в противен случай процентът на истински отрицателни резултати също се счита за добър резултат, когато показва, че имаме предвидено оплакване, което не е било регистрирано от специалния екип, както споменахме по-рано.

Случай на употреба 2: Грешка при потреблението на данни

Таблица 18 представя сравнителен анализ между действителните оплаквания на клиенти, докладвани чрез кол центъра, и прогнозите, генерирани от модела за прогнозиране на оплаквания, базиран на изкуствен интелект. Наборът от данни се състои от 16 записа, всеки от които показва дали клиентът е съобщил за проблем с услугата и дали моделът е предвидил точно това. Резултатите са категоризирани въз основа на точността на прогнозата (Истински положителен,

Истински отрицателен, Фалшиво положителен и Фалшиво отрицателен) и в съответните случаи е предоставено техническо обяснение, за да се идентифицира основната причина за прогнозирания проблем.

Table 17 прогнозни резултати с базирани на ИИ модел.

	Prediction Resultes by our AI-Based Model				
	Call Center Record	Customer Explanation	Predicted Record by our Model	Accuracy	Model Prediction and Technical Explanation
Rec#1	0		0	True Negative	
Rec#2	0		0	True Negative	
Rec#3	0		0	True Negative	
Rec#4	1	Calls disconnection	1	True Positive	Congested Network
Rec#5	0		0	True Negative	
Rec#6	0		0	True Negative	
Rec#7	0		1	False Positive	#Error
Rec#8	0		0	True Negative	
Rec#9	0		0	True Negative	
Rec#10	1	Call wait	1	True Positive	Bottlenecks Issues
Rec#11	0		0	True Negative	
Rec#12	0		0	True Negative	
Rec#13	0		0	True Negative	
Rec#14	1	Calls disconnection	1	True Positive	HSS issue (Home Subscriber Server)
Rec#15	1	Calls disconnection	0	False Negative	Provisioning Issue
Rec#16	0		0	True Negative	

➤ Истински положителни резултати (три случая):
Моделът правилно идентифицира действителните случаи на оплаквания в записи Rec#4, Rec#10 и Rec#14. Тези прогнози бяха подкрепени от технически обосновки, като например:

- Претоварена мрежа (Преп. №4)
- Проблеми с пречките (Преп. № 10)
- Проблем с HSS (домашен абонатен сървър) (Препоръка №14)

➤ Истински отрицателни резултати (11 случая):
В по-голямата част от записите, включително Rec#1, Rec#2, Rec#3 и други, моделът правилно е идентифицирал, че не е възникнало оплакване или повреда.

Тези резултати потвърждават надеждността на модела при откриване на сценарии без проблеми.

➤ Фалшиво положителен (един случай):

В Res#7 моделът неправилно е предвидил оплакване, когато от кол центъра не е докладван проблем. Този тип грешка може да доведе до ненужни предупреждения или разпределение на ресурси, въпреки че не влияе пряко на клиентското преживяване.

➤ Фалшиво отрицателен (един случай):

В Res#15 моделът не успя да открие действително оплакване, докладвано от клиента. Основната причина беше идентифицирана като проблем с осигуряването. Този тип грешка е критичен, тъй като отразява пропусната възможност за проактивно справяне с реален проблем на клиента, което потенциално може да повлияе на удовлетвореността и задържането на клиентите.

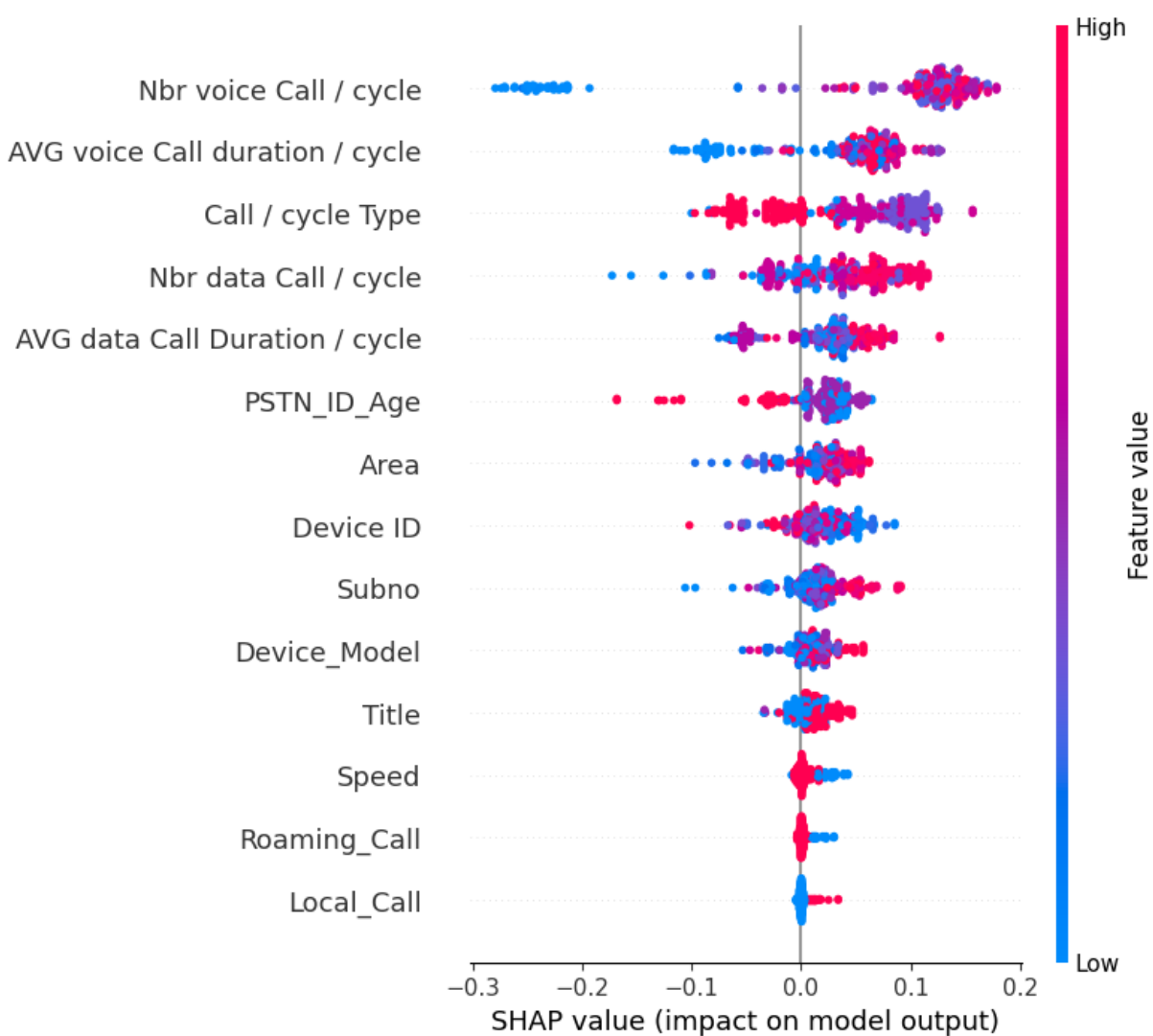
➤ Процент на точност: 87,5%

Резултатите показват, че моделът, базиран на изкуствен интелект, е ефективен при прогнозиране на оплакванията на клиентите с високо ниво на точност. Способността му да открива ключови проблеми, свързани с мрежата, като например претоварване и проблеми с конфигурацията на сървъра, подчертава неговата стойност в подкрепа на проактивното управление на мрежата. Появата на фалшиво отрицателен резултат обаче показва необходимостта от по-нататъшно усъвършенстване, особено в областта на неизправностите, свързани с осигуряването. Препоръчва се непрекъснато обучение на модела и интегриране на актуализирани данни за оплаквания, за да се подобри допълнително точността на прогнозиране и да се сведат до минимум критичните грешки.

Таблица 18 Резултати за точност за модел, базиран на изкуствен интелект

Общо записи = 16	Т. Положителен	Т. Отрицателен	Ф.Положителен	Ф. Отрицателен
	3	11	1	1
	ВЯРНО		НЕВЯРНО	
	14		2	
	Положителен %		Отрицателен %	
	0.875		0,125	

Интерпретируемост на модела с помощта на SHAP приложението:



Фигура 66 Интерпретиране на специфичните характеристики

Фигура 66 представя обобщаваща графика на SHAP, която предлага ясно разбиране на предсказуемостта на поведението на модела на ANN по отношение на оплакванията на клиентите. Ключовите изводи включват:

- **Важност и влияние на характеристиките:**
Графиката идентифицира най-влиятелните характеристики при прогнозирането на оплаквания и илюстрира как стойността на всяка характеристика (висока или ниска) влияе върху резултата от модела (към оплакване или липса на оплакване).
- **Най-важна характеристика – Брой гласови повиквания на цикъл:**
Високите стойности (червени точки) са свързани с повишена вероятност за оплаквания, докато ниските стойности (сини точки) я намаляват. Това предполага, че по-ангажираните потребители е по-вероятно да се сблъскат с проблеми или да ги съобщат.
- **Средна продължителност на гласовото повикване на цикъл:**
По подобен начин, по-дългите продължителности на повикванията са склонни да насочват прогнозата към оплаквания, докато по-кратките продължителности показват по-малко оплаквания.
- **Тип повикване на цикъл:**
Кодираните стойности, представляващи различни типове повиквания, показват, че определени типове са по-свързани с оплаквания. По-високите кодирани стойности насочват прогнозите към оплаквания.
- **По-малко влиятелни характеристики – Roaming_Call, Local_Call:**
Тези характеристики показват SHAP стойности близо до нула, което предполага минимално влияние върху прогнозирането на оплакванията.

Като цяло, графиката подобрява интерпретируемостта на модела на ИНН и предоставя приложими прозрения, особено за насочване към клиентски сегменти и усъвършенстване на стратегиите за събиране на данни.

4.6 Последици за бизнеса

В проучването е оценено поведението на клиентите при оплаквания. Според диаграмата (Фигура 13), на Етап 1, от 1000 потребители на мобилни телефони, 22,5% са се сблъскали с грешка в обслужването, докато останалите 77,5% не са се сблъскали с никакви проблеми. На Етап 2, от 225 потребители, които са се сблъскали с грешка в обслужването, 73,3% са се оплакали, показвайки недоволство от услугата, докато 26,7% са избрали да не се оплакват. На Етап 3, сред оплакващите се, 87,9% са се оплакали директно на компанията, докато 12,1% са използвали други канали, за да изразят оплакванията си, като например социални медии или потребителски форуми. На Етап 4, от тези, които са се оплакали директно на компанията, 83,4% са получили задоволително решение на проблемите си, докато 16,6% са се сблъскали с отлагане, което предполага забавяне на разрешаването или незадоволителен първоначален отговор. В резултат на това е разработен проактивен подход, използващ методи на изкуствен интелект (ИИ), за да се предвидят потенциални потребители, които не се оплакват, и други потребители, които е вероятно да имат проблеми в бъдеще. Целта беше да се предотвратят забавяния при разглеждането на оплакванията чрез своевременно предприемане на необходимите действия въз основа на прогнози от изкуствен интелект. Просто изчисление разкри, че 26,7% от клиентите, които срещат проблеми с обслужването, решават да не подават никакви оплаквания до компанията. Освен това, сред жалбоподателите, 12,1% не се оплакват директно от компанията, докато 16,6% от тези, които се свързват с

компанията, отлагат проблемите си. Част от тези случаи доведоха до отлив на клиенти. Следователно, това поражда опасения относно смекчаването на този проблем чрез използване на модели за прогнозиране, по-специално алгоритми за машинно обучение, за да се предскажат и разберат оплакванията на клиентите чрез експериментиране и сравнение.

Заклучение

През този изследователски период беше разработена цялостна рамка за машинно обучение за прогнозиране на потребителските оплаквания в телекомуникационния сектор. Проучването премина през няколко ключови фази, за да се гарантира методологична строгост и оперативна релевантност.

Първоначално беше проведено обширно събиране на данни, обхващащо регистрационни файлове за оплаквания на клиенти, показатели за качество на обслужване (напр. прекъсвания на повикванията, прекъсвания на LTE), данни за геолокация, типове устройства и модели на използване. Тези набори от данни бяха получени от взаимодействията с клиентите и записи за ефективността на мрежата.

Суровите данни преминаха през щателна предварителна обработка, за да се подобри качеството и аналитичната използваемост. Това включваше обработка на липсващи стойности, разрешаване на несъответствия, инженерни функции (като честота на прекъсвания на повикванията и тенденции в използването на данни) и кодиране на категорични променливи.

Впоследствие бяха избрани модели за машинно обучение, подходящи за класификация, включително Logistic Regression, Random Forest и изкуствени невронни мрежи (ANN). Наборът от данни беше разделен на подмножества за обучение, валидиране и тестване, за да се подпомогне разработването на модели и настройването на хиперпараметри. Беше проведен анализ на важността на характеристиките, за да се идентифицират ключови предсказващи фактори за вероятността от оплаквания.

Производителността на модела беше оценена с помощта на стандартни показатели за класификация: точност, прецизност, изчерпаемост и F1-оценка. Най-ефективният модел беше внедрен в производствена среда, което позволи прогнозиране на потенциални оплаквания в реално време. Интеграцията със системи за управление на взаимоотношенията с клиентите (CRM) и обслужване на клиенти позволи получаването на практични анализи и проактивни интервенции в обслужването.

Накрая бяха внедрени механизми за непрекъснато наблюдение и отчитане, за да се проследи ефективността на модела, да се идентифицират променящите се тенденции в оплакванията и да се информират стратегическите подобрения в обслужването. Резултатите позволиха идентифицирането на високорискови клиентски сегменти и разработката на превантивни стратегии за управление на обслужването.

Научни и приложни приноси

Това проучване има за цел да разработи предсказващи модели или алгоритми, които могат точно да прогнозират кои клиенти е вероятно да подадат бъдещи оплаквания. Това може да помогне на компаниите проактивно да решават проблемите на клиентите и да подобрят цялостното удовлетворение, като по този начин повишат бизнес стойността на своите услуги.

- 1- Идентифициране на често срещани проблеми:** Анализът на оплакванията на клиентите помага на компанията да идентифицира повтарящи се проблеми или проблемни точки на клиентите. Тази информация може да се използва за приоритизиране на подобренията в качеството на услугите или предлаганите продукти. Въз основа на този анализ е разработен модел за прогнозиране на факторите, свързани с оплакванията на клиентите.
- 2- Експериментален проект:** този модел е реализиран и тестван с помощта на различни техники за машинно обучение. Проведени са контролирани (обучени) експерименти, за да се оцени ефективността на избраните техники за намаляване на оплакванията на клиентите.

- 3- **Непрекъснато подобрене:** Получените резултати са докладвани на телекомуникационната компания. Адаптирана е система за обратна връзка, която използва анализ на оплакванията за подобряване на процесите, продуктите и услугите. Този итеративен процес осигурява непрекъснато подобряване на клиентското преживяване.

Компаниите могат да използват техники за анализ на данни и машинно обучение, за да дадат практически принос за прогнозиране на оплакванията на клиентите. Чрез прилагането на анализ на данни и алгоритми за машинно обучение, бизнесите могат проактивно да идентифицират модели и тенденции, които биха могли да доведат до недоволство на клиентите. Това се постига чрез анализ на исторически данни за оплаквания и обратна връзка от клиентите.

Резултатите от това проучване могат да бъдат използвани по следния начин:

- 1- CRM система за анализ, базирана на нови методи за анализ на оплакванията на клиентите и данните за отпадане на клиенти, за да се увеличи задържането на клиенти и да се подобри публичния образ, производителността и печалбите на компанията.
- 2- Интеграция за системата за анализ на отпадането на клиенти: В резултат на това открихме корелация между продължителността, времето и честотата на обажданията и техния ефект върху отпадането на клиенти и оплакванията. Това подчертава важността на бързото и ефективно справяне с оплакванията на клиентите, тъй като може да превърне негативните преживявания в положителни и да изгради лоялност на клиентите.
- 3- Машинно обучение и прогнозен анализ: Използването на усъвършенствани техники за анализ, като например алгоритми за машинно обучение (клъстеризиране и прогнозно моделиране), за да прогнозира поведението на клиентите въз основа на данни за оплаквания. Този проактивен подход позволява на екипа за обслужване на клиенти да предприеме превантивни мерки за намаляване на недоволството на клиентите.
- 4- Визуализация и отчитане: Използване на техники за визуализация на данни и диаграми за ясно и проникателно представяне на резултатите.

Ограничения и бъдеща работа

Това проучване е ограничено от фокуса си върху специфични типове оплаквания, данни и мрежови повреди, както и от ограничения набор от данни, обхващащ ограничен брой абонати за кратък период от време. Анализът е допълнително ограничен от разчитането на данни от една организация и изключването на алтернативни източници на данни, като например социални медии и записи на разговори.

Бъдещите изследвания трябва да включват данни за оплаквания от множество източници, да удължат периода на анализ до 6–12 месеца и да ангажират множество организации, за да се подобри обобщаемостта. Препоръчва се разработването на хибридни модели и включването на допълнителни алгоритми, като KNN и Naïve Bayes. Освен това, трябва да се проучи анализът на настроеността, използващ както контролирано, така и неконтролирано обучение, за да се обогатят прогнозните прозрения.

СПИСЪК С ПУБЛИКАЦИИ

1. Hussein Ibrahim and Vladimir Dimitrov, **Customer Behavior Analysis and Service Enhancement in Telecom Company Using Machine Learning Methods**; AIP Publishing (accepted)

2. Hussein Ibrahim, **Forecasting Customer Complaints in the Mobile Telecommunication Sectors Using Supervised Machine Learning**; AIP Publishing (accepted)

The above articles was accepted by the organizers of the 50th International Conference, Applications of Mathematics in Engineering and Economics" - Bulgaria (AMEE'24), <http://amee.fpmi.bg/>, and it will be published in American Institute of Physics (AIP) Conference Proceedings journal, which is indexed in Scopus with SJR factor (SJR2024: 0.15), Web of Knowledge (IF2024:1.4).

3.Hussein Ibrahim, Mohamad Fakherdin and Issa Kamar, **Detecting Brain Activity and Brain Disorders using Machine Learning Algorithms Applied on EEG Recordings**; Paper Published in IJSER Volume 9, Issue4, April 2021 Edition (ISSN 2229-5518), https://www.ijser.org/onlineResearchPaper_View.aspx?Detecting-Brain-Activity-and-Brain-Disorders-using-Machine-Learning-Algorithms-Applied-on-EEG-Recordings.pdf

4.Hussain Ibrahim and Mazen Hakim (2021) **Predicting customer complaints in Mobile Telecom Industry based on supervised machine learning**; JECET; June 2021-August 2021; Sec. B; Vol.10. No.3, 127-137. [DOI: 10.24214/jecet.B.10.3.12737.];http://www.jecet.org/paper_description.php?category=Comp%20Science&volume=10&issue=3