# Introduction

ge powerful tools many fields including image speech recognition natural language processing even medicine article provide review methods applications machine learning deep learning including strengths weaknesses well potential future directions also discuss challenges associated technologies including data privacy ethical considerations need transparency decision making process machine learning deep learning two revolutionary technologies field artificial intelligence become increasingly popular recent ye e security earlier several classical methods used detecting mal ware embedded various features like signature uri stic others traditional mal ware detection techniques unable defeat new generations mal ware sophisticated ob fus cation tactics deep learning increasingly used mal ware detection dl based systems per form conventional mal ware detection approaches finding new mal ware variants furthermore dl based techniques provide rapid mal ware prediction excellent detection rates analysis different mal ware types computers improve automatically experience one today rapidly growing technical fields lying intersection computer science statistics core artificial intelligence data science recent progress machine learning driven development new learning algorithms theory ongoing explosion availability online data low cost computation adoption data intensive machine learning methods found throughout science technology commerce leading evidence based decision making across many walks life including health care man uf act uri concepts techniques used solving research problem achieving research objectives indicates theoretical disciplines computer science follow machine learning techniques chopra k hur ana 202 3 machine learning sub field artificial intelligence broadly defined machine capability im itate intelligent human behavior like humans machines become capable making intelligent decisions learning past experiences machine learning employed many applications including fraud detection prevention self driving cars recommendation ri ence gain new knowledge order improve ability perform real world tasks stated liu et al 2011 researchers employed supervised learning training machine learning model predict labels based input text data text classification cai et al 2018 text classification one widely used natural language processing technologies common text classification applications include spa identification news text classification information retrieval emotion analysis intention judgment etc traditional text class ifiers based mach

# Method

n mia v villa rica month 202 4 vision laguna state polytechnic university center sustainable development initiatives transforming lives communities mission l sp u provides quality education responsive instruction distinctive research sustainable extension production services improved quality life quality policy l sp u delivers quality education responsive instruction distinctive research sustainable extension production service thus committed continual improvement meet applicable requirements provide quality e ffi st cruz campus intelligent chat bot integration collaborative features efficient research management prepared submitted sha ina may v day ape ra john je ral r del rosario kris top france pun io partial fulfillment requirements degree bachelor science computer science recommended approval acceptance \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ accepted approved partial fulfillment requirement degree bachelor science computer science date signed \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ acknowledge ments procedure performing text classification tc first needs obtained information rich term weight ing scheme achieve higher tc performance far term frequency inverse document frequency f idf widely used term weight ing scheme suffers two def iciencies first global weight ing factors idf f idf approaches infinity certain term occur text second idf equal zero certain term appears text offset draw backs first conduct depth analysis current term weight ing schemes subsequently improved term weight ing scheme called ter end develop embedded topic model et genera tive model documents marries traditional topic models word em bed ding specifically et models word cat egorical distribution whose natural parameter inner product word em bed ding em bed ding assigned topic fit et develop efficient amor tized variation al inference algorithm et discovers interpret able topics even large vo ca bular ies include rare words stop words per forms existing document models late nt dir ich let allocation terms topic quality predict ive performance abilities whereas lu et al del chat bot technology usa bility business contexts studies consideration comprehensive overview evolving landscape computational techniques various domains emerges long 202 2 mohan et al 202 3 del natural language processing nl p long emphasizes grammatical error correction mohan focuses sarcasm detection abd usa lom ov na 202 3 jai sw al sami kan nu 2017 tackle data analysis former encompassing text mining latter focusing robust ness random forest algorithm churchill singh 202 2 lu et al term relevance within document corpus method compute weight term signifies significance within document relative entire corpus f aspect derived term frequency within document divided document total term count idf component considers inverse term frequency across documents corpus two components multiplied yield term final weight researchers explored alternative term weight ing schemes enhance information retrieval text classification systems example pa ik 2013 proposed f idf variant excel capturing diverse vector ization using term frequency inverse document frequency f idf encoding process f idf vector izer employed convert prep ro ces sed text data numerical features suitable machine learning prep ro ces sed \_ texts variable contains list prep ro ces sed text strings string cleaned processed analysis f idf vector izer represented vector izer transforms text strings numerical matrix x row x corresponds prep ro ces sed text document column represents unique term found entire corpus f idf values x quan tify importance term training loop implemented deep learning model typically used tasks text classification natural language processing loop erate specified number epoch represent complete pass training data set epoch model put training mode model train within epoch loop training data divided batch es using train \_ load er batch opt imi zer gradient zero ed prepare gradient descent batch consists input \_ id \_ batch attention \_ mask \_ batch labels \_ batch model applied input data generating log raw model predictions loss function loss \_ f n cases actually turned positive defined number true positive divided number predicted values recall equation 4 recall formula source hasty 202 3 recall model would explain many actual positive cases able correctly predicted defined number true positive divided total number actual positive f1 score equation 5 f1 score formula source hasty 202 3 f1 score model would give combined idea precision recall metric would maximum precision would equal recall defined harmonic mean precision recall average precision harmonic mean precision recall average precision ap equation 6 average precision formula sum precision divided number relevant documents ranked list mean average precision map equation 7 map average precision formula mean average precision scores query q number que ries set ave p q average precision ap given query q system development methodology development research va ult advanced web based research repository laguna state polytechnic university st cruz campus intelligent chat bot integration collaborative

# Result

specialization expert sharing val ua ble knowledge providing valuable suggestions related study dr ri na j arc iga l stat istic ian providing valuable insights guidance study data sampling design ms ce zza ne dim ac ula nga n language critic assistance refining manuscript construction grammar dr jefferson l le rio associate dean college computer studies granting permission conduct study lastly researchers would like acknowledge dedication hard work member research team whose efforts crucial successful completion study hyper para meter conduct model evaluation based accuracy precision recall f1 score trained machine learning model utilized ld late nt dir ich let allocation bert bid ire ction al en code r representations transformers random forest distinct numbers parameters determine machine learning model integrate assess performance integrated models conducting actual testing within system theoretical framework theoretical framework establishes algorithms implemented system development process highlighted ja dha v et al survey ad es several pioneers industry worked steer us right direction number different algorithms one employ machine learning ml required output decides use ml algorithms characteristic ally fall one two learning types supervised un su per vis ed learning ml widely used software enable improved experience user using ml robots acquire skills learn adapt environment working robots acquire skills object placement grasping objects loco mot ion skills either automated learning learning via human intervention race ml used heal r times article cited allows users set alert receive notification new publications areas interest become widely used tool researchers students professionals seeking scholarly information conducting literature reviews academia ed u online platform allows researchers scholars academics share access scholarly articles papers research materials founded 2008 academia ed u provides space individuals create profiles load research work connect others respective fields users follow specific researchers topics om forest algorithm churchill singh 202 2 lu et al 202 2 explore computational techniques churchill singh tracing evolution topic modeling lu et al critically reviewing chat bot designs le cu n et al 2015 mo ller 202 3 introduce fundamental concepts artificial intelligence le cu n emphasizing deep learning applications mo ller providing broader understanding machine learning additionally mohan et al 202 3 churchill singh 202 2 del different aspects nl p mohan focusing sentiment analysis churchill singh highlighting

# Discussion

ng professional 3 ability apply design develop evaluate systems components processes mathematical foundations algorithm ic principles computer science theories 4 developed culture research technology advancement 5 demonstrated good leadership team player contribute nation building engage life long learning foundation professional development approval sheet thesis entitled research va ult advanced web based research repository laguna state polytechnic university st cruz campus intelligent chat bot integration c en ce date signed \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ \_ acknowledge ments researchers would like express heart felt gratitude following individuals val ua ble contributions support throughout completion study mrs mia v villa rica thesis adviser dedicated time patience valuable guidance insight ful ideas constructive criticisms greatly influenced success study mrs maria lau reen b miranda technical editor met ic ulously reviewing correct ing manuscript format content mr mark p bernardino specialization expert sharing val ua ble knowledge ed sense cam ara der ie solidarity together shared challenges triumph creating lasting memories cher ish researchers express sincere gratitude everyone involved making journey un get table experience day ape ra sha ina may v del rosario john je ral r pun io kris top france abstract research aims develop research va ult web based system efficient storage organization retrieval research materials system integrate chat bot user assistance unique feature transforming manuscripts con cise im rad format using nl p text mi manuscripts con cise im rad format using nl p text mining development research va ult involves designing system prep ro ces sing research data training machine learning models including ld random forest bert evaluating model performance conducting real world testing research design employs experimental developmental research methods system developed following sc rum software development methodology machine learning approach system development includes ld topic modeling f idf term weight ing random forest class ific ling f idf term weight ing random forest classification bid ire ction al en code r representation bert performance evaluation phase integrated algorithms tested web application assessing accuracy efficiency actual data testing success retrieval measured recall relevance retrieved documents measured precision average precision ap mean average precision map also calculated implementation research va ult promises boost research output efficiency institution crucial step towards modern izing research practices research ep towards modern izing research practices research va ult offers centralized repository research materials improving access enabling collaboration among researchers table contents pre lim ina ries page review related literature research methodology list tables list figures list equations list app end ices technical background interview communication letter forms request letter title proposal statement iso forms defense rating sheets summary recommendations curriculum vita e definition terms ensure clarity consist en um vita e definition terms ensure clarity consistency use language throughout study important define key terms concepts may different interpretations meanings section provides comprehensive list technical operational term ino logies used research along corresponding definitions explanations technical terms term ino logies used design development developed system defined section operational terms section defines terms phrases derived study operational ly implying way used study chapter introduction background res e al og ing longer sufficient includes data also analytical processes tools knowledge structures research output management stated ho uw elin g et al 202 3 essential aspect managing vast amount data information generated academic institutions must done carefully part research process ad hering fair principles using research objects ro crate solutions researchers effectively manage enhance value research outputs expanding concept research data management research output management foster ho listic approach managing r es ear ch repository system resulting research materials scattered challenging locate address challenges study proposes development advanced web based research repository system called research va ult laguna state polytechnic university st cruz campus system incorporate intelligent chat bot integration collaborative features enhance efficient research management enable seam less storage organization management research materials making easy faculty members students locate research materials quickly advanced sea r ts locate research materials quickly advanced search capabilities enable users search research materials based key words authors publication date criteria intelligent chat bot also included system assist users na vi gating research repository answering common que ries chat bot provide conversation al interface enables users access research materials insights user friendly efficient manner addition features mentioned research va ult also include special functionality allows users transform manuscript su cci nc im rad f allows users transform manuscript su cci nc im rad format using natural language processing nl p text mining techniques particular feature prove especially valuable educators students required conde nse research outcomes presentation publication purposes implementing research va ult laguna state polytechnic university st cruz campus centralized research repository system enables faculty members students quickly locate research materials collaborate efficiently thereby enhancing research output efficiency res ear c ed uce research output proposed study aims develop advanced web based research repository system called research va ult laguna state polytechnic university st cruz campus system incorporate intelligent chat bot integration collaborative features enhance research management efficiency enable seam less storage organization management research materials increase collaboration research output specifically study sought answer following research problems design develop advanced web based research repository system p advanced web based research repository system includes intelligent chat bot ri u consultation im rad convert er conde nse thesis manuscripts 10 page rs effectively prep ro ces research data including tasks text extraction pdf section classification content organization create clean structured data set machine learning model training train machine learning models including ld late nt dir ich let allocation random forest deep learning model bert bid ire ction al en code r representations transformers using prep ro ces sed r actual testing procedures research objectives general objective research develop advanced web based research repository laguna state polytechnic university st cruz campus intelligent chat bot integration collaborative features efficient research management specifically research aims design develop advanced web based research repository system chat bot integration im rad convert er laguna state polytechnic university st cruz campus allows seam less storage organization management research materials enhance e f ni zation management research materials enhance efficient research management perform data prep ro ces sing tasks including text extraction pdf section classification content organization prepare research data set machine learning model training train machine learning models using prep ro ces sed research data set leverage techniques ld late nt dir ich let allocation random forest deep learning model bert bid ire ction al en code r representations transformers various sets hyper para meter conduct model evaluation based lo pment process highlighted ja dha v et al survey 202 3 text mining face data mining del ine ated procedure di sti lling implicit knowledge textual data proving essential managing extensive volumes un st ructured text data thus framework text mining serves guide structure study figure 1 text mining framework source sa qui cel et al 2018 text mining process drawing inspiration ja dha v et al 202 3 comprehensive procedure involves several crucial stages including text prep ro ces sing discovery knowledge text prep ro ces sin rep ro ces sing discovery knowledge text prep ro ces sing initial stage deals transformation un st ructured text structured format could either document based concept based depending requirements analysis stage critical prepares raw text data making suitable processing analysis following prep ro ces sing next stage knowledge discovery document based intermediate form knowledge discovery involves uncover ing patterns trends across collection documents texts achieved application various text mining techniques techniques mm ed iate ly apparent conceptual framework proposed research repository system laguna state polytechnic university st cruz campus designed provide efficient effective platform research management knowledge sharing among faculty members students external stakeholders research va ult serve digital repository research materials accessible registered users repository offer several key features support research activities chat bot citation generator chat bot im rad convert er research va ult repository intelligent chat bo rter research va ult repository intelligent chat bot work together provide enhanced research experience users conceptual framework proposed research repository system summarized following diagram figure 2 conceptual framework study conceptual framework draws existing theories literature explain phenomenon outlines steps needed study based insights researchers viewpoint observations related study subject framework guides research process clear informed approach study integrated three essential text mining te ch study integrated three essential text mining techniques ld topic modeling f idf feature conversion random forest classification along bid ire ction al en code r representations bert process commenced ama lga mat ing various cs v data sources unified data set laying ground work streamlined data handling analytical phase sim plify ing text clusters words employing ld study revealed underlying themes within data technique illuminated prominent topics prevalence across different documents process goes towards nu meric different documents process goes towards numerical conversion transforming prep ro ces sed text weighted f idf vectors conversion allocated values based word importance within documents larger data set culminating pivotal jun cture class ifier underwent training class ifier effectively grouped text data pre de fine categories training process involved subset data set subsequent assessment occurring using independent testing set analyze effectiveness class ifier comprehensive evaluation undertaken examination pro vid hen sive evaluation undertaken examination provided insights precision recall f1 score support metric category offering clear understanding classification performance scope limitations study proposed research repository system laguna state polytechnic university st cruz campus aims provide efficient platform research management collaboration scope project limited development web based research repository well integration intelligent chat bot feature im rad convert er using nl p techniques following detailed convert er using nl p techniques following detailed scope limitations project development web based research repository incorporates key features sort filter brows ing research materials comments citation generation pdf viewer manuscript im rad convert er integration intelligent chat bot feature provides assistance answer que ries related ri u consultation proposed research repository system designed specifically laguna state polytechnic university st cruz campus may applicable institutions system rely avail abi li may applicable institutions system rely availability quality research materials uploaded users guarantee accuracy complete ness research materials significance study proposed research repository system laguna state polytechnic university st cruz campus aims provide efficient user friendly platform managing sharing research materials significance study lies potential address challenges limitations associated traditional research management approaches enhance research productivity collaboration among faculty research productivity collaboration among faculty members students specifically system output would great benefit following individuals groups research implementing unit head system would enhance ri uh efficiency providing overview student research papers enabling effective management tracking undergraduate facilitating creation im rad journal dean college system would provide dean overview progress undergraduate allowing monitor research productivity faculty members students efficiently student tor assess students research productivity efficiently researchers study would beneficial present future researchers would conducting study related text mining involves topic modeling cluster ing community system would encourage particular group researchers avoid searching research papers outside university convenience moreover institutions could replicate study communities part growing popularity digital rep osi tori es enhance research productivity collaboration chapter ii review related literature chapter pre n chapter ii review related literature chapter presents review relevant literature studies local foreign aim synth es ize assess various related sources provide comprehensive understanding topic hand review cover current research theories approaches related research repository systems chat bot integration collaborative features also explore potential benefits challenges technologies context efficient research management natural language processing nl p words long 202 2 entitled grammatical error correction model tan ding eventually affecting whole application system performance result intelligent word grammatical error detection correction english text significant difficult aspect natural language processing therefore paper examines phenomena word spelling grammatical errors undergraduate english essays balance mathematical statistical models technology solutions involved intelligent error correction research findings study represented two aspects 1 non word mistakes four sorts errors studied insertion loss replace four sorts errors studied insertion loss replacement exchange letters focuses non word mistakes varied word forms english abbreviation h yp hen ated compound terms proper nouns produced word pronunciation difficulties paper utilizes non word check information recommend opt imum combination prediction method based suggested candidate list actual word errors genuine word repair model trained approach 83 78 accurate used actual words spelling errors context 2 ri fies correct sentence grammar using context inform difficulty levels 4 6 furthermore work obtain clause correct ness rate 99 70 system average correct ive accuracy rate four level six level essays 80 furthermore gu ler ak gul 202 2 present study titled review science natural language processing highlighting today world possible normal human brain understand interpret rapidly growing data piles time use artificial intelligence techniques great requirement perception interpretation big data help machines one popular sub branches artificial intelligence natural la n ds natural language processing process begins moving words smallest building blocks make sense text digital world brings model creation process prep ro ces sing root ing feature extraction operations words transferred digital world processes connection established machine language human natural language processing technology communication human machine provided spoken language without need port input study conducted cha ud har et al 202 2 titled intelligent virtual research environment natural language process resources among nl p researchers required understanding building solution hm objective apply software engineering natural language processing concept object oriented model using collection usa ble objects defining communication protocol consisting set rules must applied exchange data two nl p modules proposed state art iv virtual environment creating modifying executing analyzing various nl p solutions technology proposed idea broadly based define iv nl p object framework model permits developer create mo object framework model permits developer create modify execute application analyze outcomes operations visual representations modules variety nl p based applications tools modules plug ins already exist publish store available environment used research community large develop complete nl p framework platform require much ass em bling collecting tools modules one place matter good tool module working individually requires standards set protocols also requires compliant composition pre defined algorithms implement lia nt composition pre defined algorithms implementation brief require comprehensive open framework bundle manage integrate set nl p tools modules components applications algorithms define associated rules comprehensive data structures knowledge recent study ali 202 2 titled ai natural language processing nl p natural language processing nl p could branch artificial intelligence ai allows machines know human language goal form systems make sense text automatically perform tasks like translation spell check topic perform tasks like translation spell check topic classification natural language processing nl p recently gained much attention representing anal ys ing human language computational ly spread applications various fields like computational linguistics email spa detection information extraction sum mar ization medical question answering etc goal natural language processing style build software system analyze understand generate languages humans use naturally could also ready address computer addressing another p machines human like facilitating easier communication two nl p witnessed significant advancement wide range applications past decades applications proven highly beneficial everyday life voice command activated machines numerous research organizations actively working enhancing nl p develop practical user friendly solutions potential nl p lies ability create computer interfaces intuitive humans enabling individuals communicate computers natural language instead requiring learn specific computer language sum requiring learn specific computer language sum studies long 202 2 gu ler ak gul 202 2 cha ud har et al 202 2 ali 202 2 san adi et al 202 2 collectively provide insights multi face ted field natural language processing nl p long 202 2 focuses practical applications nl p particularly context grammatical error correction english essays showcasing significance accurate language understanding effective human computer communication gu ler ak gul 202 2 emphasize pivotal role ai techniques including nl p processing interpreting gr ec hn iques including nl p processing interpreting growing volumes data lining nl p ability bridge gap human language machine understanding cha ud har et al 202 2 contributes proposing integrated framework iv nl p addresses need comprehensive platform un ify various nl p tools technologies foster ing research collaboration knowledge sharing nl p community ali 202 2 highlights nl p broad app lica bility computational linguistics spa detection illustrating enduring significance domain machine learning artificial int graph text representations learns complex structural semantic patterns text thereby detecting sarcastic content efficiency bert g c n compared various baseline methods noted long 202 3 entitled transfer learning sentiment classification using bid ire ction al en code r representations transformers bert model sentiment currently one emerging areas research due large amount web content coming social networking websites sentiment analysis crucial process recommend ing systems people generally purpose sentiment anal ys systems people generally purpose sentiment analysis determine author attitude toward subject overall tone document huge collection studies make effort predict useful online reviews produced conflicting results efficacy different method ologies furthermore many current solutions employ manual feature generation conventional shallow learning methods restrict general ization result goal research develop general approach using transfer learning applying bert bid ire ction al en code r representations transformers base ec tion al en code r representations transformers based model efficiency bert classification evaluated comparing similar machine learning techniques experimental evaluation proposed model demonstrated superior performance terms outstanding prediction high accuracy compared earlier research comparative tests conducted positive negative ye lp reviews reveal fine tuned bert classification performs better approaches addition observed bert class ifiers using batch size sequence length significantly affect class ific ati sequence length significantly affect classification performance explored previous literature kala iva ni et al 2020 entitled sarcasm identification detection conversion context using bert sarcasm analysis user conversion text automatic detection irony insult hurting painful ca ust ic humour vulgar ity de grade individual helpful field sentimental analysis cyber bu lly ing immense growth social media sarcasm analysis helps avoid insult hurts humour affect someone paper present traditional machine learning approach e per present traditional machine learning approaches deep learning approach l st rn n bert bid ire ction al en code r representations transformers identifying sarcasm used approaches build model identify cat ego rize much conversion context response needed sarcasm detection evaluated two social media forums twitter conversation data set red dit conversion data set cross cultural study b hard wa j et al 202 2 entitled every young ster around us uses sarcasm indirect way say negative statement growth artificial intelligence con vo lu tion al neural networks cnn long short term memory l st investigation deep 2021 entitled bid ire ction al en code r representations transformers bert language model sentiment analysis task latest trend direction sentiment analysis brought new demand understanding context ual representation language among various conventional machine learning deep learning models learning context promising candidate sentiment classification task bert new pre trained language model context em bed ding attracted attention due model context em bed ding attracted attention due deep analyzing capability valuable linguistic knowledge intermediate layer trained larger corpus fine tuned nl p task many researchers adapted bert model sentiment analysis tasks influencing original architecture get better classification accuracy article sum mar izes reviews bert architecture performance observed fine tuning different layers attention heads explored previous literature various aspects natural language processing nl p sentiment analysis mohan et ng ua ge processing nl p sentiment analysis mohan et al 202 3 focus sarcasm detection highlighting challenges posed sarcastic expressions social media texts proposing bert g c n architecture detect sarcasm accurately long 202 3 del ves sentiment analysis emphasizing importance sentiment classification web content introducing bert based model powerful approach purpose kala iva ni et al 2020 extend discussion sarcasm identification using traditional machine learning approaches l st rn n bert detect sarcasm user con vers roach es l st rn n bert detect sarcasm user conversation texts across social media platforms b hard wa j et al 202 2 address challenge sarcasm detection context artificial intelligence machine programming proposing hybrid model combines bert cnn l st efficient sarcasm detection finally deep 2021 explores use bert pre trained language model sentiment analysis discussing context ual em bed ding capabilities application improving classification accuracy collectively studies showcase versa tility nl p techniques part ic ul ud ies showcase versa tility nl p techniques particularly bert addressing various language related tasks sarcasm detection sentiment analysis focus improving accuracy efficiency understanding interpreting textual data term weight ing schemes als hri et al 202 3 presented novel supervised term weight ing scheme called f td term frequency term discrimination ability sentiment analysis tasks traditional un su per vis ed term weight ing schemes f idf may sufficient sentiment analysis text classification tasks sent ime n um ents categories called supervised term weight ing st w schemes however weigh extracted features without considering characteristics noisy features data im balance therefore study novel st w approach proposed known term frequency term discrimination ability f td f td mainly presents extracted features different degrees discrimination cat ego riz ing several groups subsequently group weighted based contribution proposed method examined four sa data set using naive bay es n b support vector machine sv models ex ai bay es n b support vector machine sv models experimental results proved superiority f td two baseline term weight ing approaches improvements ranging 0 52 3 99 f1 score statistical test results verified significant improvement obtained f td cases p value ranged 0 000 0 59 7 0 04 55 nu gr oh et al 202 2 conducted study detecting emotion indonesian wee ts te cting emotion indonesian wee ts term weight ing scheme study journal information systems engineering business intelligence 8 61 70 10 204 7 3 ji se bi 8 1 61 wo approaches assign weights features tests carried using ten fold cross validation three classification algorithms performance model measured using accuracy f1 score term weight ing schemes highest performance term frequency inverse category frequency f ic f term frequency relevance frequency f rf scheme supervised approach performed better un su per vis ed one however find consistent advantage experiments found term frequency inverse document frequency f idf also performed exceptionally well traditional f e document semantics various degrees number term weight ing schemes proposed vs improve text cat ego rization performance much evidence shows performance term weight ing scheme often varies across different text cat ego rization tasks mechanism underlying variability scheme performance remains unclear moreover existing schemes often weight term respect category locally without considering global distribution term occurrences across categories corpus paper first systematically examine pro con existing term w systematically examine pro con existing term weight ing schemes text cat ego rization explore reasons schemes sound theoretical bases chi square test information gain perform poorly empirical evaluation measuring concentration term distribute across categories corpus propose series entropy based term weight ing schemes measure distinguishing power term text cat ego rization extensive experiments five different data set proposed term weight ing schemes consistently per form state art schemes moreover find en tly per form state art schemes moreover findings shed new light choose develop effective term weight ing scheme specific text cat ego rization task dog et al 2020 proposed novel term weight ing strategy called f mono text classification effective representation relationship documents contents crucial increase classification performance text documents text classification term weight ing prep ro ces aiming represent text documents better vector space assign ing proper weights terms since calculation app rop ria f ic sd f f gm results obtained 7 different schemes show sr tf mono generally per formed schemes three data set moreover f mono promised micro f1 macro f1 results compared five bench mark term weight ing methods especially reuters 215 7 8 20 news group data set tang et al 2019 addressed def iciencies widely used term weight ing scheme f idf proposed improved scheme called term frequency inverse exponential frequency f ie f text representation necessary primary procedure performing text classification tc first e quent ly improved term weight ing scheme called term frequency inverse exponential frequency f ie f various variants proposed proposed method replaces idf new global weight ing factor ie f character ize global weight ing factor log like idf corpus greatly reduce effect feature term high local weight ing factor f term weight ing result representative feature generated carried series experiments two commonly used data sets corp ora utilizing naive bay es support vector machine class ifiers valid ate performance propose r machine class ifiers valid ate performance proposed schemes experimental results explicitly reveal proposed term weight ing schemes come better performance compared schemes sum studies tackled challenges sentiment analysis emotion detection text cat ego rization introducing innovative term weight ing approaches designed improve classification performance als hri et al 202 3 introduce f td supervised scheme per forms traditional un su per vis ed methods nu gr oh et al 202 2 explore various term weight ing schemes e al 202 2 explore various term weight ing schemes emotion detection emphasizing strengths supervised approaches acknowledging continued relevance f idf wang et al 2021 propose entropy based schemes consistently sur pass existing methods shed ding light global term distribution importance dog et al 2020 introduce f mono sr tf mono delivering superior results particularly reuters 215 7 8 20 news group data set tang et al 2019 address f idf limitations f ie f showcasing better performance text classification ow cas ing better performance text classification together papers highlight critical role tailored term weight ing enhancing text analysis classification outcomes across diverse domains text mining comprehensive review abd usa lom ov na 202 3 argued text mining science generalized language inform atics appeared basis methods machine learning rules statistics text mining also called text analytics use natural language processing nl p analyze free un st ructured text documents databases transform normal ized structured da e survey recent advances authorship tri bution text mining presented survey focuses authorship tri bution methods statistical ly computational ly supported opposed traditional literary approaches main aspects covered include changes research topics time basic feature metric machine learning techniques advantages disadvantage approach moreover corpus size number candidates data im balance result description pose challenges authorship tri bution discussed inform future work recent article ta vana et al 202 2 form future work recent article ta vana et al 202 2 review digital transformation supply chain process management using text mining industry 4 0 technologies causing paradigm shift supply chain process management digital transformation supply chains provides enormous benefits organizations em powering collaboration among multiple internal external organizations systems study presents narrative review explaining existing knowledge digital transformation supply chain process management using text mining sum mar iz hai n process management using text mining sum mar izes existing literature explain current state art supply chain digital ization comprehensive review identifies important topics technologies determines future trends emerging field investigate articles published web science sc opus databases use text mining techniques cluster ing topic modeling article contents using vo viewer bi bl iom et ric analysis 395 articles 12 700 references analyzed contents articles explored using text mining approaches synthesized result ed using text mining approaches synthesized results reveal important topics digital transformation sustainable supply chain management circular economy industry 4 0 technologies study discovers big data data analytics block chai n artificial intelligence machine learning internet things critical technologies facilitating supply chain digital transformation finally lay heat p analysis research articles found digital transformation supply chain management industry 4 0 decision making sustainability em er gin et approaches allow researchers explore large scale collections texts efficient manner given potential text mining method inquiry primary purpose manuscript enable novice experienced innovation researchers select specify document interpret text mining techniques way generates valid reliable knowledge innovation management community involved taking stock text mining applications field innovation research date means systematic review 124 journal articles employing text mining techniques published basket 10 pr ying text mining techniques published basket 10 premier innovation management 8 top general management journals results systematic manual computational analysis articles illustrate state evolution text mining applications field also allow evidence based recommendations regarding future use paper presents method ological conceptual context ual development priorities contribute establishing higher method ological standards text mining enhance method ological rich ness field mentioned literature review kumar et al es field mentioned literature review kumar et al 2015 text mining similarity search using extended tri gram algorithm reference based local repository data set emerging technological scenario world becoming digital hub data easily accessible internet take consideration academic research development fields digital resources become important academic research publications current environment easily accessed internet easy availability research work attracts academic literature dish ones ty pl agi aris many res ear demi c literature dish ones ty pl agi aris many research papers published several conference proceedings journals may percentage pl agi ari zed contents time author may cite irrelevant references research paper present research paper reference based extended tri gram approach reported check textual pl agi aris text written research papers references form pivotal part research paper dissertation defines area research state art research based original ity contribution ad ju dge present article also discusses behavior dj udged present article also discusses behavior referencing three major research categories e research papers master dissertation doctoral dissertation analysis case studies collectively shed light various aspects text mining applications different domains abd usa lom ov na 202 3 introduce text mining science rooted machine learning statistical rules emphasizing role transforming un st ructured text structured data suitable artificial intelligence machine learning algorithms zheng et al 202 3 del authorship g algorithms zheng et al 202 3 del authorship tri bution text mining focusing statistical ly computational ly supported methods addressing challenges like data im balance result description context ta vana et al 202 2 explore digital transformation supply chain management text mining identifying key topics technologies shaping field including sustainability industry 4 0 technologies anton et al 2020 conducts systematic review text mining applications innovation research offering insights state evolution text n research offering insights state evolution text mining domain providing recommendations future use kumar et al 2015 tackles issue textual pl agi aris academic research proposing extended tri gram approach reference based similarity checking highlighting importance maintaining original ity proper referencing research papers dissertation together papers showcase versa tility growing relevance text mining various research domains authorship tri bution supply chain management pl agi aris detection high li g intelligence become increasingly popular recent years due ability make predictions analyze large data set provide insights previously impossible obtain article explore basics machine learning deep learning differences applications impact various industries machine learning deep learning transforming way interact technology unlock ing new possibilities innovation technologies already made significant impacts various industries potential continue revolution ize world article provides comprehensive overview bas world article provides comprehensive overview basics machine learning deep learning differences applications impact society focus current literature research article aims provide better understanding potential machine learning deep learning implications future mathew et al 2021 present study deep learning techniques overview deep learning class machine learning performs much better un st ructured data deep learning techniques per form ing current machine learning techniques enables computational models learn ni ng techniques enables computational models learn features progressively data multiple levels popularity deep learning amplified amount data available increased well advancement hardware provides powerful computers article comprises evolution deep learning various approaches deep learning architecture deep learning methods applications shin de et al 2018 august present study review machine learning deep learning applications machine learning one fields modern computing world plenty research undertaken make rn computing world plenty research undertaken make machines intelligent learning natural human behavior made essential aspect machines well various techniques devised traditional machine learning algorithms applied many application areas researchers put many efforts improve accuracy machine lea rn ing algorithms another dimension given thought leads deep learning concept deep learning subset machine learning far applications deep learning explored definitely going cater solving issues several new application ni tion gen omics el uc ida te back pro pa gation facilitates adjustment internal parameters representation computation layer meanwhile sharif ani et al 202 3 provide comprehensive review machine learning deep learning el uc ida ting strengths weaknesses applications future prospects emphasize revolutionary potential technologies highlighting impact diverse industries addressing associated challenges like data privacy ethical considerations mathew et al 2021 offers overview deep learning emphasizing superiority handling ie w deep learning emphasizing superiority handling un st ructured data evolution approaches architecture methods applications shin de et al 2018 explore broad landscape machine learning deep learning applications highlighting evolution potential address various challenges across domains lastly go pina th et al 202 3 focus deep learning based mal ware detection techniques highlighting effectiveness combat ing evolving threats providing rapid mal ware prediction excellent detection rates thus offering insights cr uc ia lent detection rates thus offering insights crucial domain cyber se cu rity machine learning ml course study mo ller 202 3 entitled machine learning deep learning states machine learning sub category artificial intelligence enabling computers ability pattern recognition continuously learn making predictions based data carry decisions without specifically programmed context machine learning broader category algorithms able use data set identify patterns discover insights enhance understanding make decisions pre di ns ight enhance understanding make decisions predictions compared machine learning deep learning particular branch machine learning makes use machine learning functionality moves beyond capabilities deep learning algorithm interpreted layered structure tries replicate structure human brain capabilities enable machine learning deep learning algorithms usage applications identify respond cyber cr imi nal manifold cyber att ack achieved analyzing big data set cyber se cu rity incidents identify patterns malicious ac um ans machines become capable making intelligent decisions learning past experiences machine learning employed many applications including fraud detection prevention self driving cars recommendation systems facial recognition technology intelligent computing book helps begin ners learn art science machine learning presents real world examples leverage popular python machine learning ecosystem topics covered book include machine learning basics supervised un su per vis ed learning linear regression log istic reg learning via human intervention race ml used healthcare analytics number start ups looking advantages using ml big data provide healthcare professionals better informed data enable make better decisions study gupta et al 202 2 entitled machine learning techniques digital library services discusses digital library organised system digital storage digital computing communications mechanism data software needed reproduce em ulate execute services point digit ization need digit ization digit ization process cha llen ion need digit ization digit ization process challenges digit ization hardware software digit ization challenges library professional role library professionals digital library initiatives future digital libraries india artificial intelligence machine learning technique help enhance quality digital library services end users context study jordan mitchell 2015 entitled machine learning trends perspectives prospects states machine learning addresses question build computers improve automatically experience one chopra k hur ana 202 3 emphasize machine learning ability replicate human intelligence learning experiences applications various domains bell 202 2 del ves machine learning algorithms distinguishing supervised un su per vis ed learning highlighting wide ranging applications including robotics healthcare analytics gupta et al 202 2 discusses role machine learning enhancing digital library services emphasizing potential improving user experiences managing digital resources lastly jordan mitchell 2015 provide broader ces lastly jordan mitchell 2015 provide broader perspective noting rapid growth machine learning driven evolving algorithms increasing data availability widespread adoption across diverse fields emphasizing shift toward evidence based decision making random forest context study jai sw al j k sami kan nu r 2017 february entitled application random forest algorithm feature subset selection classification regression disc ards insignificant variables produces efficient improved prediction performance class var ia b performance paper take gan der improvements random forest history till date approach take recorded view improvement prominently effective classification procedure begin history random forest main technique proposed br eim successful applications utilized random forest finally comparison class ifiers paper proposed give non specialists simple access principle thoughts random forest noted abstract mo ha pa tra et al 2020 entitled mo ha pa tra n sh rey k chin may 2020 optimization algorithms implemented making field optimization algorithms implemented making field machine learning efficient comparing various solutions opt imum satisfactory answer found yield better accuracy score earlier existing one paper optimization random forest performed supervised learning model classification regression detailed analysis optimization technique model done follows une qual weight voting strategy weight assigned based well individual tree performs according report ren et al 2017 research machine learning framework based random fore arch machine learning framework based random forest algorithm continuous development machine learning industry academia released lot machine learning framework based distributed computing platform widely used however existing framework machine learning limited limitations machine learning algorithm choice parameters interference noises high using threshold paper introduces research background machine learning framework combined commonly used random forest algorithm machine learning classification algorithm l gor ith machine learning classification algorithm puts forward research objectives content proposes improved adaptive random forest algorithm referred ar f basis ar f designs implements machine learning framework examination dish et al 202 2 entitled performance analysis machine learning models intrusion detection system using gin imp urity based weighted random forest gi wr f feature selection technique protect network resources sensitive data intrusion detection system id become fundamental component organ ion system id become fundamental component organizations prevents cyber cr imi nal activities several approaches introduced implemented th wart malicious activities far due effectiveness machine learning ml methods proposed approach applied several ml models intrusion detection system order evaluate performance models un sw n b 15 network ton \_ io data set used line analysis data set comparatively newer ns l k dd data set represent modern day attacks however performance analysis carried training testing decision rm ance analysis carried training testing decision tree dt gradient boost ing tree gb multi layer per ce pt ron ml p ada bo ost long short term memory l st gate rec urrent unit gr u binary classification task performance id deter ior ates high dimensional feature vector opt imum set features selected gin imp urity based weighted random forest gi wr f model embedded feature selection technique technique employed gin imp urity splitting criterion trees adjusted weights two different classes im balance data make learning different classes im balance data make learning algorithm understand class distribution based upon importance score 20 features selected un sw n b 15 10 features network ton \_ io data set experimental result revealed dt performed well feature selection technique trained models experiment moreover proposed gi wr f dt per formed existing methods surveyed literature terms f1 score sum jai sw al sami kan nu 2017 demonstrate random forest aids feature subset selection classification regression improving prediction per f lass ification regression improving prediction performance handling issues like missing data lier noisy data effectively parma r et al 2019 provide overview random forest evolution highlighting history main technique proposed br eim successful applications comparisons class ifiers emphasizing high accuracy effectiveness mo ha pa tra et al 2020 del optimization random forest focusing supervised learning model classification regression particularly exploring une qual weight voting strategies based individual election technique enhance performance compare various machine learning models intrusion detection highlighting effectiveness gi wr f dt model achieving superior results especially terms f1 score compared existing methods topic modeling churchill singh 202 2 conducted comprehensive analysis evolution topic modeling study titled evolution topic modeling topic models applied everything books newspapers social media posts effort identify prevalent themes text corpus provide depth analysis un su per vis ed topic model us provide depth analysis un su per vis ed topic models inception today trace origins different types contemporary topic models beginning 1990s compare proposed algorithms well different evaluation approaches throughout also describe settings topic models worked well areas new research needed setting stage next generation topic models study titled review topic modeling methods va yan sky kumar 2020 topic modeling popular analytical tool evaluating data numerous methods topic modeling developed consider many kinds tho ds topic modeling developed consider many kinds relationships restrictions within data set however methods frequently employed instead many researchers gr avi tate late nt dir ich let analysis although flexible adaptive always suited modeling complex data relationships present different topic modeling approaches capable dealing correlation topics changes topics time well ability handle short texts encountered social media sparse text data also briefly review algorithms used opt imi ze fer parameters topic mod l gor ith ms used opt imi ze fer parameters topic modeling essential producing meaningful results regardless method believe review encourage diversity performing topic modeling help determine topic modeling method best suits user needs die ng et al 2020 present study topic modeling em bed ding spaces topic modeling analyze documents learn meaningful patterns words however existing topic models fail learn interpret able topics working large heavy tailed vo ca bular ies end develop embedded topic model et genera tive cat ion terms topic quality predict ive performance bard e bain wad 2017 conducted overview topic modeling methods tools study titled overview topic modeling methods tools topic modeling powerful technique analysis huge collection document topic modeling used discovering hidden structure collection document topic viewed recurring pattern co occurring words topic includes group words often occurs together topic modeling link words context differentiate across uses words different meanings paper discuss methods words different meanings paper discuss methods topic modeling includes vector space model vs late nt semantic index ing l si pro ba bilis tic late nt semantic analysis pl sa late nt dir ich let allocation ld features limitations discuss tools available topic modeling gen si stand ford topic modeling tool box mall et big art applications topic modeling covered topic models wide range applications like tag recommendation text cat ego rization key word extraction information filtering similarity search fields text mining n filtering similarity search fields text mining information retrieval hu et al 2014 conducted study titled interactive topic modeling topic models useful ubiquitous tool understanding large corp ora however topic models perfect many users computational social science digital humanities information studies machine learning experts existing models framework often take leave proposition paper presents mechanism giving users voice encoding users feedback topic models correlation words topic model framework models correlation words topic model framework interactive topic modeling allows un train ed users en code feedback easily erative ly topic models late ncy interactive systems crucial develop efficient inference algorithms tree based topic models valid ate framework simulated real users discussed research papers collectively provide extensive view evolution diversity applications topic modeling field natural language processing text analysis churchill singh 202 2 offer comprehensive analysis historical dev e 202 2 offer comprehensive analysis historical development topic models examining origins algorithm ic variations evaluation approaches setting stage future topic modeling va yan sky kumar 2020 emphasize need broader selection topic modeling methods beyond late nt dir ich let analysis introducing various approaches capable handling complex data relationships temporal changes topics short text data die ng et al 2020 introduces embedded topic model et enhance traditional topic modeling word em bed ding improving int iona l topic modeling word em bed ding improving interpret ability even large vo ca bular ies per form ing existing models bard e bain wad 2017 provide overview popular topic modeling methods tools including vector space model late nt semantic index ing pro ba bilis tic late nt semantic analysis late nt dir ich let allocation along practical applications text mining information retrieval lastly hu et al 2014 introduce interactive topic modeling enabling non expert users incorporate feedback topic models enhancing user incorporate feedback topic models enhancing user friend liness efficiency topic modeling applications chat bot study lu et al 202 2 titled critical review state art chat bot designs applications chat bots intelligent conversation al agents interact users natural languages chat bots perform variety tasks many companies committed numerous resources develop deploy chat bots enhance various business processes however lack date critical review thoroughly examines state art technologies innovative applications chat bots fully realize business value chat bots assist making sensible decisions related development deployment chat bots various business contexts researchers interested design development chat bots also gain useful insights critical review identify fruit ful research topics future research directions based research gaps discussed study long 2021 entitled development intelligent telegram chat bot using natural language processing intelligent chat bots gaining interest past years due advance artificial intel li gen er est past years due advance artificial intelligence algorithms result many studies conducted emotional transition dial og structures one benefits medical applications psychological assessment clinical counseling autism diagnostic advanced cognitive models could provided hand chat bot knowledge generally comes web based information repository information reliable rather versatile contain emotions objective paper develop intelligent chat bot using natural language processing telegram api first text processing age processing telegram api first text processing using telegram api python developed next emotion recognition performed recorded chat appropriate response sent user results showed telegram chat bot could interact smoothly users identify user emotions study intelligent chat bot lab security automation prasad et al 2020 artificial intelligence based com mun icative artefacts called chat bots purpose chatter bot chat bot render interaction human robot form speech text offer best services variety areas education x offer best services variety areas education healthcare transportation etc per research nearly 85 product offerings automated 2020 proposed system voice based chat bot helps improve security automation lab system makes use automatic speaker recognition r algorithm order recognize person allow inside lab allows authorized person access lab facilities algorithm used generating list available components lab based person key word input thus auto mat ing lab components dispatch outlined po ong od et al 2019 chat one nts dispatch outlined po ong od et al 2019 chat bot computer code interact humans without taking human assistance chat bot uses natural language processing nl p understanding human request query blog information network people group post information area expertise people similar interests read learn problem traditional blog websites user get data linked home page also user manually go every page get data proposed research work aims designing building chat bot administrators common man interact english ask bot administrators common man interact english ask questions like show posts abc hence provides natural language nl interface user language english blog information network database study sat u et al 2015 entitled review integrated applications aim l based chat bot artificial intelligence mark language aim l derived ex tens ible mark language xml used build conversation al agent chat bot artificial ly developed lot works make conversation al agent low cost configuration availability make possible use various app l ratio n availability make possible use various applications paper give brief review applications used aim l chat bot conversation al service applications related cultural heritage e learning e government web base model dial og model semantic analysis framework interaction framework humor ist expert network management adaptive modular architecture well case providing useful services also interact customers give solution problems aim l chat bot instead human beings popular day day entrepreneur users provide efficient pu lar day day entrepreneur users provide efficient service summary lu et al 202 2 offer comprehensive review chat bot technologies applications across different business sectors emphasizing need br id ging research gaps proposing future directions academia industry long 2021 focuses development intelligent chat bot emotional recognition particularly medical applications showcasing potential chat bots psychological assessment counseling prasad et al 2020 present ai based chat bot designed lab security automation based chat bot designed lab security automation demonstrating versa tility chat bots enhancing security stream lining lab processes po ong od et al 2019 discuss use chat bots context blog networks enabling users interact naturally english retrieve specific information blog databases sat u et al 2015 provide insights aim l based chat bots integration various applications highlighting cost effectiveness con fi gur ability chat bot technology software development model study um eu go et al 202 3 investigates effect differ en ware requires sm es majority software publishers adopt practice ss dl c promoting ss dl c need know efforts adapted various sd lc models study empirical ly examined effect sd lc models innovation characteristics ss dl c derived diffusion innovation theory intention adopt ss dl c sample software security managers software sm es united states surveyed sd lc model used perception relative advantage trial ability ob ser va bility complexity compatibility ss dl c intention adopt ss dl c k rus kal wallis test performed data showed stat lc k rus kal wallis test performed data showed statistical ly significant differences sd lc model groups relative advantage compatibility trial ability ob ser va bility complexity intention adopt ss dl c results also indicated sm e software security managers average would inclined adopt ss dl c given impetus ss dl c adoption efforts mostly uniformly applied across sd lc models software security policy makers may find results study useful ss dl c adoption policy formulation patel 202 3 provides insight software development life el 202 3 provides insight software development life cycle sd lc process models paper highlights software development industry rapidly grown recent past years let look sectors like healthcare tech education e commerce transport food beverages etc depend software developed software main part grown industries businesses whether form websites ai based machines robots apps computer software saves time regarding bits helps solve complex problems prolonged repeated processes fast accurately without error handle pro b roc esses fast accurately without error handle problems features industries need software programs facilitate employees administration work offices banks departments etc developing suitable high quality software according requirements given client primary goal software engineering software requires plan whole team team members cover different parts software software project manager team follow specific sd lc software development life cycle model throughout project achieve goals complete project time according je ct achieve goals complete project time according given requirements sd lc models covered depth paper kara 202 3 focuses using secure sd lc models develop secure web applications software takes part center digital transformation also co vid 19 accelerated digital transformation every organization tries transfer processes manuel electronic environment provided software transformation also enforce corporate use secure software development life cycle sd lc methodology framework web applications special internet wo log framework web applications special internet work application great importance digital transformation due reasons accessible anywhere world requiring special client application study investigated web application security requirements met sd lc al zai k hal fan 202 2 examine involvement top management different phases system development life cycle sd lc one essential factors success system implementation recognized top management support involvement research however addressed question sort engagement necessary study investigates function top management various phases system implementation help us understanding support mechanism top management various sd lc stages achieve goal author performed qu ali tative study five different firms kuwait interviewing top management project management system analysts managers research established criteria top management participation indicated top management involved primarily planning implementation phases well phases needed gupta 2021 presents comparative study different sd lc mod 2021 presents comparative study different sd lc models software development life cycle sd lc refers methodology clearly defined processes creating high qual ity software cost effective reliable method software developing process quite systematic structural sd lc defines framework different activities tasks administered software development process software development process quite complex without proper planning would ef fi cie nt use sd lc models make software development process simple systematic various soft wa lo pment process simple systematic various software development life cycle models used software development process advantages limitations paper included six sd lc models waterfall model spiral model v model agile model erative model rapid application development ra model um eu go et al 202 3 examine effect different sd lc models adoption secure sd lc ss dl c innovation characteristics highlighting uniform ity ss dl c adoption efforts across various sd lc models patel 202 3 provides overview sd lc process models em ph el 202 3 provides overview sd lc process models emphasizing importance rapid growth software development industry role facilitating development high quality software kara 202 3 focuses use secure sd lc models developing secure web applications emphasizing significance web applications digital transformation al zai k hal fan 202 2 investigate top management involvement different phases sd lc lining importance top management support involvement ensuring success information system projects gupta 2021 offers com pa n form ation system projects gupta 2021 offers comparative study different sd lc models lining advantages limitations models waterfall spiral agile erative rapid application development ra related systems table 1 features comparison existing systems research gate social networking platform founded 2008 designed specifically scientists researchers academics collaborate share scientific research provides digital space researchers create profiles share publications engage discussions connect peers users lo ions engage discussions connect peers users load full text versions papers join research groups follow researchers stay updated latest developments field research gate become popular platform scientists di sse minate work increase visibility foster collaborations aiming facilitate knowledge exchange communication within scientific community beyond however also faced legal challenges criticism concerning copyright infringement sharing unauthorized full text articles google scholar web based search engine pr articles google scholar web based search engine provided google focuses specifically scholarly literature including articles books conference papers prep rin ts launched 2004 google scholar aims help researchers academics find relevant scholarly information wide range disciplines index es content various sources including academic publishers universities websites making comprehensive resource academic research google scholar provides citation metric number times article cited allows users set alert fields users follow specific researchers topics interest enabling receive updates new publications research activities platform also offers features networking discussion forums ability track citations analytics uploaded papers academia ed u aims foster collaboration facilitate dissemination knowledge within academic community providing platform researchers showcase work connect wider audience however important note content academia ed u may require subscription subject copyright restrictions synthesis discus subject copyright restrictions synthesis discussed long 202 2 mohan et al 202 3 del natural language processing nl p focus different aspects long 202 2 discusses grammatical error correction emphasizing importance accurate text analysis nl p addresses non word mistakes spelling errors training word repair models contrast mohan et al 202 3 concentrate sarcasm detection using advanced techniques like bid ire ction al en code r representations transformers bert graph con vo lu tion al networks g c n studies involve nl p lo con vo lu tion al networks g c n studies involve nl p long work centered error correction mohan paper explores complex ities identifying sarcasm text showcasing versa tility nl p similarly abd usa lom ov na 202 3 jai sw al sami kan nu 2017 focus data analysis different applications methods abd usa lom ov na 202 3 discuss text mining broad approach utilizes natural language processing statistics convert un st ructured text structured data emphasize applications various domains including missing data imp utation contrast jai sw al sami k ing missing data imp utation contrast jai sw al sami kan nu 2017 concentrate random forest algorithm highlighting role feature selection data analysis discuss robust ness especially handling noisy data papers address data analysis abd usa lom ov na explore broader field text mining jai sw al sami kan nu del deeply specific algorithm capabilities related vein churchill singh 202 2 lu et al 202 2 explore applications computational techniques different contexts churchill singh 202 2 provide analysis evolution topic modeling g sentiment analysis churchill singh highlighting topic modeling historical development abd usa lom ov na 202 3 le cu n et al 2015 touch artificial intelligence abd usa lom ov na discussing text mining versa tility le cu n et al emphasizing deep learning applications jai sw al sami kan nu 2017 lu et al 202 2 explore computational techniques data analysis jai sw al sami kan nu focusing random forest algorithm capabilities lu et al examining chat bot technology usa bility business contexts reviewed studies collectively span wide ra text reviewed studies collectively span wide range topics fields natural language processing nl p sentiment analysis text mining deep learning machine learning random forest topic modeling chat bot technology software development life cycle sd lc models studies showcase multi face ted nature nl p focus practical applications data interpretation collaborative platforms also addressing sentiment analysis challenges emphasizing role tailored term weight ing deep learning machine learning explored extensively demon ar ning machine learning explored extensively demonstrating transform ative impact various domains including cyber se cu rity studies random forest highlight evolution optimization application intrusion detection context topic modeling various approaches tools discussed handle complex data relationships additionally chat bot technologies applications across sectors examined sd lc models evaluated impact secure software development chapter iii research methodology chapter outlines methodology employed achieve ob je c ha pt er outlines methodology employed achieve objectives study presents research design local studies well data collection methods data model generation development methodology performance evaluation algorithms software used system architecture software testing research design research design research va ult advanced web based research repository laguna state polytechnic university st cruz campus intelligent chat bot integration collaborative features efficient research management utilize experimental develop research management utilize experimental developmental research design methods stated yn ye mi 202 3 experimental research scientific approach research one independent variables manipulated applied one dependent variables measure effect latter effect independent variables dependent variables usually observed recorded time aid researchers drawing reasonable conclusion regarding relationship 2 variable types method utilized order determine best performing trained natural language processing nl p algorithm model ed natural language processing nl p algorithm model terms extract ing information respect data set collected hand developmental research defined rich ey 1994 systematic study designing developing evaluating instructional programs processes products must meet criteria internal consistency effectiveness involves situation product development process analyzed described final product evaluated therefore researchers used method examine consistency efficacy development system local e study study conducted laguna stat ec tion prevention self driving cars recommendation systems facial recognition technology intelligent computing figure 3 classification machine learning techniques source sur ya kan thi 2020 machine learning enable computers ability pattern recognition continuously learn making predictions based data carry decisions without specifically programmed stated mo ller 202 3 researchers study would use ml automatically classify sections within pdf document extract content predictions determine text segment ed sections fa predictions determine text segment ed sections facilitating conversion pdf im rad format natural language processing techniques ali 202 2 natural language processing nl p could branch artificial intelligence ai allows machines know human language goal form systems make sense text automatically perform tasks like translation spell check topic classification natural language processing nl p recently gained much attention representing anal ys ing human language computational ly spread applications various fields li k ds smallest building blocks make sense text digital world explained gu ler ak gul 202 2 nl p used researchers handle various aspects text processing classification making possible automatically cat ego rize structure content pdf document based sections un su per vis ed learning aka nk sha \_ rai et al 202 3 supervised learning name indicates presence supervisor teacher basically supervised learning teach train machine using data well labelled means data already tagged correct answer machine provided new set examples data rec answer machine provided new set examples data supervised learning algorithm analyses training data set training examples produces correct outcome labeled data figure 5 un su per vis ed learning source mori moto pont 2021 un su per vis ed learning algorithms allow users perform complex processing tasks compared supervised learning although un su per vis ed learning unpredictable compared natural learning methods un su per vis ed learning algorithms include cluster ing anomaly detection neural networks discussed johnson mal detection neural networks discussed johnson 202 3 researchers employed un su per vis ed learning specifically late nt dir ich let allocation ld discover hidden topics within collection text documents data modeling process supervised learning cunningham et al 2008 supervised learning accounts lot research activity machine learning many supervised learning techniques found application processing multimedia content defining characteristic supervised learning availability ann ota ted training data name vo kes id ng ines email filtering etc text classification difficult task due high dimensional feature vector comprising noisy irrelevant features various feature reduction methods proposed eliminating irrelevant features well reducing dimension feature vector relevant reduced feature vector used machine learning model better classification results explored ag ar wal mit tal 2014 researchers used text classification cat ego rize text documents four im rad sections introduction method result discussion algorithm analysis sect n method result discussion algorithm analysis section algorithm analysis presents method ologies techniques used researchers applying machine learning development system en ume rates ld topic modeling f idf term weight ing scheme random forest class ifier bid ire ction al en code r representation bert used achieve objectives study also includes overview different machine learning algorithms architecture advantages disadvantage provide better understanding general behavior used development system ld topic model behavior used development system ld topic modeling introduction methodology results discussion ld vital component data prep ro ces sing pipeline aiding extraction relevant topics text employing ld identify cat ego rize thematic content section helping us distinguish introductory content method ological descriptions result discussions analytical interpretations figure 8 ld plate diagram source et al 2020 advantages ld include pro ba bilis tic model interpret able topics needing know topics advance however di sa va 3 proposed f idf variant excel capturing diverse aspects term importance making effective short long que ries additionally chen et al 2016 f gm scheme incorporating novel statistical model demonstrated superior performance compared traditional f idf supervised term weight ing schemes text classification studies provide valuable insights enhancing design development information retrieval text classification systems im rad section classification often involves distinguishing distinct sections enable e ffi cie es distinguishing distinct sections enable efficient information extraction analysis applied task f idf ability weigh importance terms within document corpus becomes particularly valuable instance introduction section certain terms phrases may indicative background information methods section might contain specialized technical terminology utilizing f idf assign appropriate weights terms allowing model disc ern unique linguistic characteristics im rad section improves accuracy classification also enhance idf variant adept capturing diverse aspects term sal ien cy catering short long que ries additionally chen et al 2016 developed f gm scheme incorporating novel statistical model demonstrated superior performance text classification compared traditional f idf supervised term weight ing schemes im rad section classification complex task demands ability disc ern distinct linguistic patterns content structures within section random forest ability construct multiple decision trees combine outputs makes particular l e decision trees combine outputs makes particularly well suited task im rad section exhibit distinct linguistic patterns vocabulary random forest algorithm excel capturing nu ances randomly selecting subset features specific terms phrases decision point within trees algorithm effectively disc ern unique characteristics section example methods section might frequently include technical terms experimental details distinguish general language found introduction analytical results presented results section lever na ly tical results presented results section lever aging random forest model efficiently learn differentiate linguistic cues leading accurate reliable classification results im rad section ultimately enhancing automation efficiency scientific document analysis organization bid ire ction al en code r representation figure 10 overall structure bert model source sun et al 202 2 bert bid ire ction al en code r representations transformers powerful model greatly advanced natural language processing nl p text classification di al language processing nl p text classification distinctive feature capability understand word context left right sides training text classification using bert involves two key phases pre training model learns language representations extensive text corp ora fine tuning bert adapted specific tasks additional output layer trained labeled data study published ieee transactions geo sc ience remote sensing et al 2020 hs bert hyper sp ect ral image bid ire ction al en code r representations transformers introduced hyper sp e e bert bid ire ction al en code r representations transformers represents significant advancement field natural language processing nl p greatly enhance accuracy precision classification tasks bert distinguishing feature unique bid ire ction al context understanding training enabling capture complex linguistic nu ances intricate relationships words im rad section classification means model effectively disc ern specific key words also broader context appear leading robust context aware classification process lever aging b context aware classification process lever aging bert pre tuning model acquire deep understanding linguistic patterns structural characteristics define im rad section making highly adapt able capable handling wide range scientific documents varying writing styles content structures enhanced context ual understanding em power model make accurate nu ance predictions ultimately improving automation efficiency scientific document analysis organization data collection methods researchers used data collection method methods researchers used data collection methods interviews online research collection data information critical study researchers conducted short interviews library services head director college computer studies cc gain access research materials collected manuscripts served primary source training machine learning models used system table 2 data set manuscript im rad interview table 3 document titles document pages data set data model generation section would en ume rate various method ologies employed de would en ume rate various method ologies employed developing algorithm models would include data prep ro ces sing token ization feature extraction topic modeling training transfer learning based models data prep ro ces sing data prep ro ces sing crucial step text mining tasks involves transforming raw text data manage able form machine learning algorithms process achieved various techniques converting characters lower case removing stop words pun ct uation token ization stemming le mma ti zation vector ization techniques help ing text \_ chunk collection chunks code employs token izer segment text token utilizes return \_ tensor parameter return token format compatible p yt ch popular deep learning library additionally pad ding tr un cation parameters used ensure token uniformly sized enabling efficient processing machine learning models token ization fundamental prep ro ces sing step text analysis enabling transformation un st ructured text data structured format readily consumed various natural language processing tasks text class ific natural language processing tasks text classification sentiment analysis machine translation feature extraction feature extraction process extract ing relevant features raw data natural language processing involve techniques bag words term frequency inverse document frequency f idf word em bed ding feature extraction important step preparing text data machine learning algorithms figure 13 feature extraction using f idf figure demonstrates essential step text vector ization using term frequency inverse doc um es extraction late nt topics prep ro ces sed text making valuable tasks document cat ego rization content understanding training training refers process teaching machine learning algorithm make predictions decisions based data natural language processing involve training algorithms decision trees support vector machines neural networks text data training involves adjusting parameters algorithm minimize error training data set figure 15 pre tuning bert model figure training loop implemented deep learning model ty pi log raw model predictions loss function loss \_ f n employed compute difference model predictions ground truth labels cumulative loss epoch updated back pro pa gation performed adjust model parameters using gradient descent opt imi zer step processing batch es within epoch total loss epoch printed code ni ppet ex em pl ifies training process deep learning models goal mini mi zing loss improving model predict ive capabilities data model evaluation order calculate performance transfer learning based bid ire ction al en code r nc e transfer learning based bid ire ction al en code r representation models would need run test set test set used evaluation im rad data set due fact testing developed transfer learning based bid ire ction al en code r representation models well would return performance levels since models used prediction section content metric classification algorithms based confusion matrix used figure 16 confusion matrix confusion matrix crucial tool evaluating performance bert based text classification model confusion matrix used ed text classification model confusion matrix used quan tify model classification results breaking predictions four categories true positive true negative false positive false negative breakdown allows detailed analysis model behavior helping identify specific areas excel may making errors example reveal whether model tends confuse certain classes performs well across classes information valuable fine tuning model making decisions threshold values gaining insights strengths weaknesses ultimately aiding g insights strengths weaknesses ultimately aiding ref ine ment classification task metric derived confusion matrix used measure performance models discussed following accuracy equation 2 accuracy formula source hasty 202 3 accuracy model would measure often class ifier correctly predict defined ratio number correct predictions total number predictions precision equation 3 precision formula source hasty 202 3 precision model would explain many correctly predicted cases actually turned positive defined number tr mp us intelligent chat bot integration collaborative features efficient research management system followed software development life cycle sd lc model specifically sc rum software development methodology sc rum agile framework managing completing complex projects based principles transparency inspection adaptation sc rum methodology involves several key stages including product back log sprint planning sprint back log sprint inc rem ent sprint review sprint retrospective figure 17 sc rum software development method ol ti figure 17 sc rum software development methodology source sc rum 202 3 product back log sc rum work flow starts construction product back log simply list features functional ities system product items list chosen based suggestions stakeholders ideas development team items also prior iti zed high priority items top list non priority items bottom product back log regularly maintained ensure prior iti zation correct new items added removed needed table 4 product back log researchers sprint planning sprint planning dev el researchers sprint planning sprint planning development team decides product back log items worked upcoming sprint meeting held determine delivered achieved duration sprint also decided meeting sprint back log output sprint planning sprint back log list product back log items worked sprint table 5 sprint back log first sprint sprint sprint development team works completing items sprint back log daily sc rum meetings held discuss progress ensure everything track inc rem ent outcome sprint usable product called inc cr ement outcome sprint usable product called inc rem ent integration completed items sprint back log sprint review end sprint sprint review held development team presents inc rem ent stakeholders feedback sprint retrospective sprint review sprint retrospective held development team reflects performance sprint identifies areas improvement performance evaluation algorithms performance evaluation phase researchers assessed effectiveness algorithms integrated web application including ld topic model random forest b li cation including ld topic model random forest bid ire ction al en code r representation bert done actual testing data allowing objective evaluation models performance terms accuracy efficiency relevant metric ld topic modeling evaluation ld topic model primarily based quantitative metric designed assess quality topics generated model metric provide objective measures model performance without need human interpretation process begins data prep ro ces sing text data multiple cs v files undergoes essential task data multiple cs v files undergoes essential tasks token ization stemming named entity recognition missing values also handled appropriately following prep ro ces sing code employs late nt dir ich let allocation ld topic modeling uncover ing late nt themes within text corpus text data transformed numerical features using f idf vector ization preparing machine learning performance evaluated using dedicated test data set generates classification report offering essential metric like accuracy precision recall f1 sc ia l metric like accuracy precision recall f1 score assessing model performance ensure model availability future use saved file using job lib library allowing make predictions new unseen text data needed figure 18 classification report ld model training classification report provided figure based performance ld late nt dir ich let allocation model classify ing im rad sections offers insights model ability correctly classify different section types precision measures often model positive predictions correct rec ur es often model positive predictions correct recall assess es ability capture relevant instances f1 score balanced metric combines precision recall context ld model demonstrates moderate high performance classify ing im rad sections overall accuracy 88 precision recall values vary across sections method discussion sections exhibiting higher precision recall compared introduction result sections metric collectively indicate model effectiveness cat ego riz ing im rad sections room improvement areas random forest ra sections room improvement areas random forest random forest algorithm utilized perform text classification model performance evaluated using common metric accuracy precision recall f1 score evaluation conducted test data set split training set testing set training set used train random forest model training phase model learns patterns relationships within prep ro ces sed text data make predictions testing set employed assess model performance serves independent data set evaluate well trained model general iz dent data set evaluate well trained model general ize new unseen text documents evaluation results provide insights model accuracy effectiveness classify ing documents pre de fine categories figure 19 classification report random forest model training classification report presented figure sum mar izes performance random forest model classify ing im rad sections provides essential evaluation metric class including precision recall f1 score assess model ability make correct predictions section type high precision v correct predictions section type high precision values indicate model generally makes accurate positive predictions section meanwhile recall values show model effectiveness identifying relevant instances class f1 scores strike balance precision recall offering overall measure classification performance context model achieve high level accuracy correctly classify ing im rad sections accuracy 89 metric collectively demonstrate model capability classify im rad sections effectively bid ire ction al en code r rep res e sections effectively bid ire ction al en code r representation bert bert algorithm underwent comprehensive evaluation employing range performance metric including accuracy precision recall f1 score evaluation process involved utilization dedicated test data set thoughtfully divided training set validation set training set designated primary purpose training bert model allowing learn available data meanwhile validation set reserved assessing model performance ensuring rigorous evaluation results evaluation ex hi bi ring rigorous evaluation results evaluation exhibited bert model com men dable performance across multiple dimensions notably terms accuracy lining efficacy classify ing text data according specified im rad labels researchers conducted depth analysis attention weights within bert model scrutiny attention weights provided valuable insights inner workings model shed ding light makes predictions insights enhanced researchers understanding model also paved way potential improvements refining overall performance ot ential improvements refining overall performance ho listic evaluation approach sco res thorough ness rig research effort figure 20 epoch loss training bert model output provided figure represents training progress bert model classify ing im rad sections five epoch loss values steadily decrease 26 83 52 first epoch 9 07 35 fifth epoch trend indicates model progressively improving predict ive accuracy learns training data lower loss values sign ify better alignment model predictions actual im rad section lab e g n ment model predictions actual im rad section labels demonstrating model effectively adapting parameters task figure 21 confusion matrix training bert model confusion matrix presented sum mar izes performance bert model im rad section classification consists 4 x 4 matrix row corresponds true labels column represents predicted labels made model values matrix denote number instances falling specific combinations true predicted labels instance top left cell 52 signifies count accurate predictions true label intro du fi es count accurate predictions true label introduction model correctly classified conversely matrix also highlights mis class ification bottom right cell 38 indicating count instances true label discussion model predicted discussion figure 22 classification report bert model training classification report figure offers comprehensive evaluation bert model performance im rad section classification provides precision recall f1 score metric im rad section class introduction method result discussion overall sum e recall 1 00 suggests model correctly identified instances introduction data set f1 score 0 96 balanced measure precision recall macro average weighted average metric provide overall sum mar ies model performance across classes showing model average f1 score 0 93 indicating strong overall classification performance achieved accuracy 0 93 reflecting proportion correctly classified instances data set metric collectively indicate bert model performed well im rad section classification software used variety soft w section classification software used variety software employed development testing algorithms research following sections provide discussion software used study visual studio code visual studio code source code editor developed microsoft windows linux mac os includes support de bu gging embedded gi control gi th ub syntax highlighting intelligent code completion ni ppet code ref act orin g highly custom iza ble large active community contributes development provides support users python libraries several python li b support users python libraries several python libraries used research include os provides portable way using operating system dependent functionality allows interaction underlying operating system including file directory management process management environment variable manipulation secrets used generating crypt ographic ally strong random numbers suitable managing data password account authentication security token related secrets provides secure way generating random data suitable use crypt ographic app era ting random data suitable use crypt ographic applications spa cy open source software library advanced natural language processing provides tools tasks token ization part speech tag ging named entity recognition dependency par sing text classification io provides python interfaces stream handling allows creation manipulation streams data including reading writing files torch open source machine learning library python provides tensor computation strong gp u acceleration widely used field deep learning provides el eration widely used field deep learning provides tools building training neural networks p yp df 2 pure python library built pdf tool kit capable extract ing document information title author etc splitting documents page page merging documents page page crop ping pages merging multiple pages single page en cr yp ting dec ry pt ing pdf files provides wide range functionality working pdf files report lab open source engine creating complex data driven pdf documents custom vector graphics provides tools generating high q vector graphics provides tools generating high quality pdf documents rich format ting layout options transformers state art natural language processing library tensor flow 2 0 p yt ch includes pre trained models bert bid ire ction al en code r representations transformers fine tuned specific tasks text classification named entity recognition panda open source data analysis manipulation tool built top python programming language provides data structures data frame series allow easy manipulation tab ular data sk lea series allow easy manipulation tab ular data sk lea rn metric provides several functions measure quality predictions classification algorithms including classification \_ report function generates text report showing main classification metric confusion \_ matrix function compute confusion matrix evaluate accuracy classification fl ask micro web framework written python allows build web applications quickly minimal code provides tools routing request handling make easy develop web applications database connection easy develop web applications database connection database connection established using db con ne ction class db \_ connection module allowed efficient storage retrieval data used research class provides methods connecting database server executing sql que ries managing transactions system architecture section presents system architecture developed system en ume rate functions features highlight integration computing science solutions discuss structure behavior views system users interact research va ult system web system users interact research va ult system web application using front end client developed using html cs java script front end provides user friendly interface brows ing research materials leaving comments generating citations utilizing platform features communicate back end server via http requests api fetch display research data back end research va ult system responsible processing user requests managing data implementing core functionality platform built using fl ask framework python handle http requests p sing fl ask framework python handle http requests provide routing different features back end server interact sql ite database retrieve store research materials user data comments relevant information system database uses sql ite store research materials user profiles comments metadata related research papers lightweight database system suitable smaller scale applications efficiently handle data storage retrieval research va ult additionally research va ult includes im rad convert er created using bert bid ire ction al im rad convert er created using bert bid ire ction al en code r representations transformers nl p natural language processing techniques random forest model convert er analyze structures research papers according im rad introduction methods results discussion format converted documents stored sql ite database easy retrieval access software testing software testing defined saw ant et al 2012 critical process aimed evaluating software identify rec tify errors research context testing focused assessing accuracy perform text testing focused assessing accuracy performance integrated models designed classify ing im rad sections researchers conducted testing system development particular emphasis evaluating functionality web application assess performance im rad convert er researchers made use comprehensive data set containing diverse research manuscript materials data set served input system testing phase encompassing wide range research papers reflecting real world variations academic writing styles disciplines content data set ad emi c writing styles disciplines content data set fed system testing system utilized various machine learning specifically researchers harness ed power text classification techniques cat ego rize sections research manuscripts distinct im rad structure components introduction methods results discussion researchers also employed bid ire ction al en code r representation transform er bert state art transform er based model gain deep understanding context ual nu ances relationships within text advanced technique sign ific ions hips within text advanced technique significantly improved accuracy im rad section classification researchers also utilized random forest algorithm enhance overall classification accuracy involved ag gre gating predictions multiple decision trees contributing system proficiency im rad classification late nt dir ich let allocation ld topic modeling played vital role uncover ing underlying themes topics within research manuscripts added layer analysis refined classification process assess system performance com p ss ification process assess system performance comprehensive ly researchers employed several well established evaluation metric confusion matrix allowed visual ize models performance helping identify true positive true negative false positive false negative monitoring training process multiple epoch provided insights quickly models converge whether training necessary classification reports offered detailed statistics precision recall f1 score support im rad section providing thorough understanding models ad section providing thorough understanding models proficiency classify ing research manuscripts