Plagiarism Checker X - Report

Originality Assessment

1%

Overall Similarity

Date: Jan 12, 2025 (02:01 PM)
Matches: 13 / 2460 words
Sources: 1

Remarks: Low similarity detected, consider making necessary changes if needed.

Verify Report:
Scan this QR Code

AIE425IntelligentRecommenderSystems,FallSemester24/25

PersonalizedEducationalGameRecommendationSystem

221100583,MohammedHaithamMohammed 221100500,MaryamEslamSameer

2.Introduction: ThePersonalizedEducationalGameRecommendationSystemisa groundbreakingapplicationdesignedtobridgethegapbetweentechnology, education,andentertainment.Intoday’sdigitalage,educationalgameshave becomeavitalresourceforfosteringcognitivegrowth,enhancingproblemsolvingabilities,andengagin glearnersofallages.However,thesheer volumeandvarietyofeducationalgamesavailableoftenmakeitdifficultfor userstoidentifygamesthatalignwiththeirinterests,educationalgoals,and agegroup.Thisprojectaddressesthischallengebyintroducinga personalizedrecommendationsystemthattailorsgamesuggestionsto individualneedsandpreferences.

Thissystemisbuiltonrobustmachinelearningalgorithmsandsemanticdata analysistechniquestoensureprecisionandrelevanceinrecommendations.It leveragesawell-curateddatasetenrichedwithmetadata,includinggenres, descriptions,ratings,andagerequirements.Bypreprocessingandanalyzing thisdata,thesystemcreatespersonalizedrecommendationsthatarenotonly engagingbutalsoeducationallybeneficial.Designedforstudents,educators, parents,anddevelopers,thesystemprovidesauser-friendlyinterfacethat simplifiestheprocessofdiscoveringmeaningfuleducationalgames. 2.1.SystemOverview: ThePersonalizedEducationalGameRecommendationSystemoperates asaspecializedenginedesignedexclusivelyfortheeducationalgaming niche.Unlikegeneral-purposerecommendationengines,thissystem focusesonthedualgoalsofengagementandpedagogy.Byanalyzing userinputssuchasfavoritegenres,previouslyplayedgames,andage group,thesystemprovidesrecommendationsthatbalanceentertainment valuewitheducationalobjectives.

Atitscore,thesystemisdrivenbyarichdatasetcomprisingdetailed attributesforeachgame.Theseattributesincludegamedescriptions, genreclassifications,ageratings,userreviews,andothermetadata. Advancedpreprocessingtechniquesensurethatthisdataisclean, consistent,andactionable.Forinstance,thesystemclusterssimilargame titles,removesnoisefromtextualdescriptions,andencodesuser preferencestofacilitateefficientcomputation.

Thesystemintegratesmultiplealgorithms,includingcollaborativefiltering, K-meansclustering,andsemanticanalysisusingTF-IDF.These algorithmsworktogethertoidentifypatternsinuserpreferences,group similargames,andextractmeaningfulinsightsfromtextualdescriptions. Bycombiningthesetechniques,thesystemdeliverspersonalized recommendationstailoredtoeachuser’suniqueprofile. Age-appropriatenessisakeyfeatureofthesystem.Usersundertheage of15receiverecommendationsforgamessuitablefortheirdevelopmental stage,whileolderusersareofferedabroaderselectionofeducational tools.Thisfeatureensuresthatthesystemmeetstheneedsofitsdiverse audiencewhilemaintainingtrustandreliability. 2.2.ObjectivesoftheSystem: Thesystemhasbeendesignedwithclearandmeasurableobjectivesto addressthechallengesfacedbyusersindiscoveringrelevanteducational games:

Personalization:Thesystemaimstoprovidehighlytailoredgame recommendationsbyanalyzinguserinputssuchasgameplayhistory, favoritegenres,anddemographicinformation.Thisensuresthateach userreceivessuggestionsthatalignwiththeiruniquepreferencesand goals.
EducationalRelevance:Acoreobjectiveofthesystemistoprioritize gameswithstrongpedagogicalvalue.Theseincludegamesdesignedto teachskillssuchasproblem-solving,teamwork,programming,orliteracy.
AgeAppropriateness:Thesystemenforcesstrictage-basedfiltering, ensuringthatyoungerusersareexposedtocontentsuitablefortheir cognitiveandemotionaldevelopment. AdvancedTechnology:Byemployingstate-of-the-artalgorithms,the systemenhancestheaccuracyandrelevanceofitsrecommendations. Techniquessuchascollaborativefiltering,clustering,andsemantic analysisareusedtoidentifymeaningfulrelationshipsbetweenuser preferencesandgameattributes.
ScalabilityandAdaptability:Thesystemisbuilttohandlelargedatasets andintegratefutureexpansions,suchasnewdatasourcesoremerging algorithms,ensuringlong-termrelevanceandusability. Theseobjectivesguidethedesignandimplementationofthesystem, ensuringitservesawiderangeofuserswhilemaintainingafocuson personalizationandeducationalvalue. 2.3.PracticalApplicationDomain: Educationalgamesrepresentauniqueintersectionoftechnologyand learning,combiningtheimmersivequalitiesoftraditionalgameswith

structurededucationalobjectives.Thesystemfocusesspecificallyonthis domain,makingitavaluabletoolforavarietyofusecases. Educationalgamescatertodiverseaudiences,fromyounglearners developingfoundationalskillstoolderusersseekingadvancedknowledge. Forexample:

- ForYoungLearners:Gamesdesignedforusersunder15oftenfocuson teachingbasicliteracy,numeracy,andlogicalreasoning.Thesegames aretypicallyvisuallyengagingandinteractive,fosteringaloveforlearning.
- ForOlderUsers:Advancededucationalgamestargetskillssuchas teamwork,criticalthinking,andprogramming.Thesegamesoften simulatereal-worldscenarios,suchasfinancialmanagementorscientific research,makingthemidealforolderstudentsandprofessionals.
- ForInstitutions:Schoolsanduniversitiesincreasinglyusegamified learningtoolstoenhancestudentengagementandoutcomes.This systemhelpseducatorsidentifygamesthatalignwiththeircurriculaand teachingobjectives.

Byaddressingtheneedsoftheseaudiences,thesystemnotonly enhancesindividuallearningexperiencesbutalsocontributestothe broaderadoptionofeducationaltechnology. 2.4.StakeholderExpectations: ThePersonalizedEducationalGameRecommendationSystemis designedtomeetthediverseneedsofitsstakeholders,eachofwhomhas distinctexpectations:

- StudentsandLearners:Expectengagingandpersonalized recommendationsthatalignwiththeirinterestsandeducationalgoals.
- ParentsandEducators:Relyonthesystemtoprioritizeage-appropriate contentandeducationalvalue,ensuringthatrecommendedgamesare bothsafeandbeneficial.
- GameDevelopers:Benefitfrominsightsintouserpreferences,allowing themtocreatetargetedandsuccessfulproducts.
- EducationalInstitutions:Usethesystemtoidentifyhigh-quality educationalgamesthatcomplementtheirteachingmethodsandcurricula.

Tomeettheseexpectations,thesystemincorporatesadvanced algorithmsanduser-friendlydesignprinciples.Forexample,itprovides detailedexplanationsforeachrecommendation,helpingusersunderstand whyspecificgamesaresuggested. 2.5.SystemDesignandConstraints: Thesystem’sarchitecturehasbeencarefullycraftedtobalancetechnical sophisticationwithpracticalusability.Itintegratesmultiplecomponents, eachdesignedtoaddressspecificchallengesintheeducationalgaming domain. 2.5.1.DataModel: Thedatasetservesasthebackboneofthesystem,containingrichmetadata foreachgame,including: GameDescriptions:ProcessedusingNLPtechniquestoextract semanticinformation. GenreClassifications:Encodedtofacilitateclusteringandfiltering. UserRatingsandReviews:Usedtogaugethepopularityandqualityof games. AgeRequirements:Enforcedthroughstrictfilteringmechanisms. Preprocessingsteps,suchasclusteringsimilargamesandcleaningtextual data,ensurethatthedatasetisconsistentandactionable. 2.5.2.AlgorithmicFramework: Therecommendationengineintegratesmultiplealgorithmstoenhance accuracyandrelevance: CollaborativeFiltering:Identifiespatternsinuserpreferencesand interactions. Clustering(K-means):Groupsgameswithsimilarattributes,ensuring thatrecommendationscaptureshared characteristics. SemanticAnalysis(TF-IDF):Extractsmeaningfulrelationshipsfrom gamedescriptions,aligningrecommendationswithuserpreferences. Thesealgorithmsworkinharmonytodeliverrecommendationsthatare personalized,educational,andcontextuallyrelevant. 2.5.3.Constraints: Age-BasedFiltering:Recommendationsaretailoredtotheuser’sage grouptoensuresuitability. EducationalFocus:Gameswithoutclearpedagogicalobjectivesare excludedfromrecommendations.

DataQuality:Thesystem’sperformancereliesonthecompletenessand accuracyofthedataset,emphasizingtheimportanceofrigorous preprocessing.

2.6.SystemFeaturesandComponents: Component Description Dataset Includesmetadatasuchasgenres,ageratings,descriptions,anduser reviews. Preprocessing Techniques Textcleaning,clusteringofgameversions,andsemanticanalysisvia TF-IDF. Recommendation Methods Collaborativefiltering,clustering(K-means),andage-basedfiltering. UserInputs Agegroup,favoritegenres,previouslyplayedgames,andlearning preferences. Outputs Rankedlistofeducationalgameswithdetailedexplanationsfor recommendations. 3.DataCollectionandPreprocessing: ThePersonalizedEducationalGameRecommendationSystemdependson high-quality,structureddatatoproducemeaningfulandrelevant recommendations.Datacollectionandpreprocessingformthecornerstoneof thisproject,transformingrawinformationintoactionableinsightsthatpower thesystem’salgorithms.Thesestepsarenotmerelypreparatory—theyare integraltothesystem'sperformance,accuracy,andabilitytodeliver personalizedsuggestionsthatmeetuserpreferencesandconstraints. Thedatasetusedforthisprojectprovidesmetadatafor980games,each describedthroughavarietyofattributesthatcapturetheirunique characteristics.Thismetadataofferscriticalinsightsintotheeducationalvalue, gameplaystyle,anduserexperienceofeachgame,enablingthesystemto alignitsrecommendationswithspecificuserneeds.However,aswithmost datasets,therawdatawasincomplete,inconsistent,andscatteredacross multiplesources,necessitatingarigorousandmulti-steppreprocessing pipeline.3.1.DataCollection: Datacollectionisthefirstandperhapsthemostcriticalstageinbuildinga recommendationsystem.Forthisproject,thedatasetwassourcedfroma publicrepositorycontainingrichmetadataaboutgamesspanningvarious genresandplatforms.Thisdatasetwasselectedforits comprehensiveness,coveringessentialfieldssuchasgamenames, genres,descriptions,ratings,agerequirements,andsupportedlanguages.

However,whilethedatasetprovidedastrongfoundation,itrequired significantenhancementstomeetthesystem’srequirementsfor personalizationandaccuracy. Thedatasetwascomposedofstructuredfieldsthatcapturedboth quantitativeandqualitativeaspectsofgames.TheNamecolumnserved astheuniqueidentifierforeachgame,whiletheShortDescriptioncolumn providedabriefnarrativeaboutthegame’scontentandobjectives.Other criticalfieldsincludedRequiredAge,whichensuredage-appropriate

recommendations,andRatings,whichprovidedanumericalmeasureof usersatisfaction.TheGenrescolumnslistedthegame’sclassifications, suchas“Educational,”“Action,”or“Adventure,”whiletheSupported Languagescolumnindicatedthelanguagesinwhichthegamewas available,offeringanadditionallayeroffilteringbasedonuser preferences.

3.2.ChallengesinDataCollection:

Despiteitsrichness,thedatasethadseverallimitations.Manyentries wereincomplete,withmissingvaluesinfieldssuchasShortDescription orGenres.Somegamesappearedmultipletimesunderdifferenteditions orversions,creatingredundanciesthatcoulddistorttherecommendation process.Additionally,thedatasetlackedinformationaboutcertainfields, suchaspromotionalimagesanddetailedgenrehierarchies,whichwere deemedcriticalforenhancingtheuserexperience.

Toaddressthesechallenges,supplementarydatawasgatheredfrom reputablesecondarysources,includingofficialgamewebsites,online gamingdatabaseslikeSteamandIGDB,andgamedeveloperpages. Theseadditionalsourceswereusedtofillmissingvalues,enrichexisting fields,andvalidatetheaccuracyofcriticalattributes.Forinstance,where Genreswereincompleteorambiguous,externaldatabaseswere referencedtoprovideamoredetailedclassification.Missingdescriptions weresupplementedwithsummariesscrapedfromofficialwebsites, ensuringthateverygamehadsufficientcontextualinformationforanalysis.

3.3.DataPreprocessing:

Thepreprocessingphasetransformedtheraw,inconsistentdatasetintoa clean,structuredformatsuitableformachinelearningalgorithms.This stepinvolvedacombinationofcleaning,enrichment,transformation,and featureengineeringtoensurethedata'susabilityandrelevance.Each stageofpreprocessingwasdesignedtoaddressaspecificissueinthe dataset,frommissingvaluestoinconsistentnamingconventions.

3.4.CleaningandNormalization:

Cleaningthedatasetwasthefirststepinpreprocessing,focusingon removinginconsistenciesandstandardizingthedata.Missingvaluesin criticalfieldssuchasRatingsandRequiredAgewereaddressedusing domain-specificimputationstrategies.Forexample,missingratingswere filledwiththemedianratingofgameswithinthesamegenre,ensuringthat theimputedvaluesdidnotdistorttheoveralldistribution.Similarly, missingdescriptionswerereplacedwithplaceholderssuchas"No descriptionavailable,"whichcouldbeexcludedfromdownstreamanalysis. Textnormalizationwasappliedtoalltext-basedfields,includingName, ShortDescription,andGenres.Thisinvolvedconvertingalltextto lowercase,removingspecialcharacters,andtrimmingexcessive whitespacetoensureuniformityacrossentries.Thesestepsreduced variabilityinthedataandimprovedtheperformanceofalgorithmsthatrely ontext-basedfeatures.

Duplicateentrieswereanothersignificantchallenge,particularlyfor gamesreleasedinmultipleeditionsorwithslightvariationsintheirtitles. Usingacombinationofclusteringandmanualverification,duplicaterows wereidentifiedandmergedintosingle,unifiedentries.Thisprocess ensuredthateachgamewasrepresenteduniquelyinthedataset,

eliminatingredundanciesthatcouldskewtherecommendations. 3.5.SemanticEnrichment:

OneofthemostvaluablefieldsinthedatasetwastheShortDescription,

whichprovidedanarrativeoverviewofeachgame.Tomakethese

descriptionsusableformachinelearningalgorithms,theywere

transformedintonumericalrepresentationsusingTermFrequency-Inverse

DocumentFrequency(TF-IDF).Thismethodquantifiedtheimportanceof

wordsineachdescriptionrelativetotheentiredataset,enablingthe

systemtoidentifysemanticsimilaritiesbetweengames.Forexample,a

gamedescribedas“Learnmaththroughengagingpuzzles”wouldbe

semanticallylinkedtoothereducationalgameswithsimilarthemes.

TreducethedimensionalityoftheTF-IDFvectorsandenhance

computationalefficiency,SingularValueDecomposition(SVD)was

applied.Thistechniqueretainedthemostrelevantfeaturesofeachvector

whilediscardingnoise,ensuringthatthesemanticrepresentationofeach

gamewasbothaccurateandcompact. 3.6.UnificationandGrouping:

Gametitlesoftenappearedinmultipleformsduetovariationsinnaming

conventions.Forexample,“7DaystoDie”and“7DaystoDieDeluxe

Edition”weretreatedasseparateentriesintherawdataset.Toaddress

this,atitleunificationprocesswasimplementedusingacombinationof

clusteringalgorithmsandmanualcuration.Titleswereclusteredbasedon

theirsemanticsimilarity,andthemostrepresentativenamewasselected

astheunifiedtitleforeachgroup.Thisensuredthatthedatasetaccurately

reflectedtheuniqueidentitiesofthegames. 3.7.GenreEncoding:

TheGenresfieldswereamongthemostcomplexattributesinthedataset,

withuptosevenseparatecolumnsforeachgame’sclassifications.To

simplifyprocessing,thesefieldswereconsolidatedintoasingle

categoricalfeature,witheachgenrerepresentedasabinarycolumn.This onehotencodingapproachallowedthesystemtoidentifygamesthat spannedmultiplegenres,suchas“Educational”and“Adventure,”without losinggranularity.Missinggenrevalueswerereplacedwith“Unspecified,” ensuringthateverygamecouldbeincludedintheanalysis. 3.8.FinalDatasetTransformation: Afterpreprocessing,thedatasetwastransformedintoastructuredformat optimizedforrecommendationalgorithms.Thetablebelowillustratesa sampleofthefinaldataset: Name TF-IDF Vectors (Sample)IsFree Short DescriptionSupported LanguagesHeader Image Website Ratings Genres 7Daysto Die [0.59,0.35,-0.40]FALSEAnopenworldgame combining first-person shooter, survival horror,and RPG elements.English, French, German, Spanish Link http://www.7daystodie.com 7 Multiplayer, Action, Adventure

Page9of50 ADance ofFire andIce[0.32,0.38,0.68]FALSEArhythm game whereyou guidetwo planets alonga winding path. English, Spanish, Korean Link https://7thbe.at 6 Indie APlague Tale: Innocence[0.52,0.51,0.20]FALSEFollowthe taleof Amiciaand Hugo throughthe darkest hoursof history. English, French, German Link https://www.focushome.com/games/a-plague-taleinnocence 7 Singleplayer, Action, Adventure AStory AboutMy Uncle [0.62, 0.20,-0.09]FALSE A platforming adventure aboutaboy searching forhis uncle. English, French, German Linkhttp://gonenorthgames.com/games/astory-about-my-uncle/ 6 Singleplayer, Adventure, Indie AWayOut [0.48,0.33,0.41]FALSE A cooperative prison escape game. English, French, German Linkhttps://www.ea.com/games/a-way-out 7 Multiplayer, Action, Adventure Thisstructureddatasetpowerstherecommendationenginebyproviding clean,enriched,andactionableinformationabouteachgame.

3.9.RoleofPreprocessinginRecommendations:

Thepreprocesseddatasetisthebackboneoftherecommendationsystem, enablingitto:

1.AccuratelycalculatesemanticsimilaritiesbetweengamesusingTF-IDF vectors.

2.Groupgameswithsharedcharacteristicsthroughclusteringandgenre encoding.

3.Filterrecommendationsbyuser-specificcriteria,suchasage,language, andpricingpreferences.

Bytransformingrawdataintostructuredinsights,preprocessingensuresthat thesystemdeliversaccurate,relevant,anduser-centricrecommendations.

4.DatasetDescription:

ThedatasetpoweringthePersonalizedEducationalGameRecommendation Systemcomprisesmetadatafor980games.Thismetadataisstructured acrossmultiplecolumns,eachprovidingcriticalinformationaboutthegames' features,accessibility,anduserexperience.Thedatasethasbeencarefully processedtoensureitsupportstherecommendationsystem'sobjectiveof deliveringpersonalized,relevant,andeducationalgamerecommendations.

4.1.DetailedDescriptionoftheDataset: Thedatasetcontainsthefollowingcolumns:

Name:Theuniquetitleofeachgame,servingasitsprimaryidentifier.

RequiredAge:Theminimumagerecommendedforplayingthegame, ensuringage-appropriatesuggestions. IsFree:Abinaryindicatorspecifyingwhetherthegameisfreetoplay.

ShortDescription:Aconcisesummaryofthegame’scontent,objectives, andgameplaystyle.

SupportedLanguages:Alistoflanguagesinwhichthegameisavailable.

HeaderImage:Linkstopromotionalimagesthatrepresentthegame visually.

Website:URLslinkingtoofficialgamepagesorstores.

Ratings:Anumericalscorereflectingthequalityorpopularityofthegame.

Categories:Ageneralclassificationofthegame,suchas"Multiplayer"or "Singleplayer."

Genres(genre1togenre7):Uptosevengenreclassifications,capturing thegame'smultifacetedcharacteristics.

4.2.ModelingUserInterests,Interactions,andIntentions:

Thedatasethasbeenstructuredtomodeluserpreferencesandbehaviors throughitsvariouscolumns.Thismodelingenablesthesystemtodeliver personalizedrecommendationsbyleveragingthefollowingfactors: 4.2.1.UserInterests:

UserinterestsareprimarilycapturedthroughtheCategoriesandGenres columns.Forinstance:

Usersinterestedinmultiplayerexperiencesaredirectedtowardgames

labeledas"Multiplayer"underCategories. Themulti-

genreclassification(genre1togenre7)supports

recommendationsforuserswhoenjoycomplexordiversegameplay

styles,suchascombining"Educational"with"Simulation." 4.2.2.UserInteractions:

Thedatasetallowsthesystemtoanalyzepatternsingamepreferencesbased

onsharedattributeslikegenresandratings.Usingclusteringalgorithms,

gameswithoverlappingfeaturesaregroupedtogethertoensure

recommendationsaligncloselywithuserpreferences. 4.2.3.UserIntentions:

TheRequiredAgecolumnensuresthatrecommendationsalignwithuserage

groups,whileIsFreeenablesbudget-conscioususerstoreceivefree-to-play

options.Theseexplicitandimplicitsignalsrefinetherecommendationprocess

tomatchuserexpectations. 4.3.ExplanationofDatasetColumns:

Thedatasetcolumnsaredetailedbelow,outliningtheirpurposeandrole

intherecommendationsystem: 4.3.1.Name:

Thiscolumncontainsthegametitles,ensuringeachgameisuniquely

identifiedinthedataset.Duplicatetitles,suchaseditionsorspecialreleases,

werestandardizedduringpreprocessingtomaintainclarity. 4.3.2.RequiredAge:

Theminimumrecommendedageforplayingeachgame.Thisfieldfilters

recommendationstoensureusersbelow15yearsoldreceiveage-appropriate

suggestions,whileolderusersarepresentedwithunrestrictedoptions. 4.3.3.IsFree:

Abinaryfield(TrueorFalse)indicatingwhetheragameisfreetoplay.This

columnallowsthesystemtofilterrecommendationsbasedontheuser's budgetpreferences.

4.3.4.ShortDescription:

Thiscolumnprovidesaconcisesummaryofthegame’scontent.Forexample:

"ADanceofFireandIce"isdescribedas"Astrictrhythmgamefocusing ontiming."

ThesedescriptionswereprocessedusingTF-IDFvectorizationto calculatesemanticsimilaritiesbetweengames,enablingthesystemto groupandrecommendgameswithoverlappingthemes. 4.3.5.SupportedLanguage: Thiscolumnlistsalllanguagessupportedbythegame,ensuringaccessibility foradiverseuserbase.Forexample,agameavailableinEnglish,Spanish, andKoreanisprioritizedforuserswiththeselanguagepreferences. 4.3.6.HeaderImage: Linkstopromotionalimagesenhancetheuserinterfaceofthe recommendationsystem,offeringavisuallyengagingwaytopresentgame suggestions. 4.3.7.Website: ThiscolumnprovidesURLstoofficialgamepages,allowinguserstoexplore moredetails,reviews,orpurchaseoptionsfortherecommendedgames. 4.3.8.Ratings: Thenumericalratingofeachgame,typicallyonascaleof1to10.Higherratedgamesaregivenpriorit yinrecommendations,reflectinguser satisfactionandquality. 4.3.9.Categories: Abroadclassificationofthegame,suchas"Multiplayer"or"Singleplayer." Thisfieldisusedtoalignrecommendationswithuserpreferencesforsocialor sologameplay. 4.3.10.Genres(genre1togenre7): Themulti-genreclassificationcapturesthediversenatureofgames.These fieldsallowthesystemtorecommendgamesthatalignwithuser-selected genres.Forexample,agamecategorizedunder"Educational,""Adventure," and"Simulation"cancatertousersinterestedinacombinationofthese genres.4.4.SummaryofDatasetColumns: ColumnName Description RoleintheSystem Name Uniquegametitle. Ensures distinct representationofgames. RequiredAge Minimumrecommendedage.Filters age-appropriate games. IsFree Indicateswhetherthegameis free. Alignsrecommendationswith userbudgetconstraints. ShortDescription Summaryofgamecontent.Enablessemanticsimilarity analysis. SupportedLanguagesListslanguagesinwhichthe gameisavailable. Filtersgamesbasedon languagepreferences. HeaderImage Linktoapromotionalimage.Enhances the recommendationinterface

Page12of50 visually. Website OfficialgamepageURL. Directsuserstoadditional gameinformation. Ratings Numericalratingfromusers.Prioritizeshigh-qualitygames inrecommendations. Categories Generalclassification(e.g., Multiplayer). Matchesuserpreferencesfor socialorsologameplay. Genres (genre1– genre7) Multi-categoryclassification ofgames. Alignsrecommendationswith user-selectedinterests.

5.DataAnalysisandInsights:

Thissectiondelvesintotheanalysisofthedataset,focusingoncritical attributessuchasrequiredage,ratings,genres,andclustering.Theinsights derivedfromtheseanalysesguidethedesignandimplementationofthe PersonalizedEducationalGameRecommendationSystem.Detailed visualizationsarepresentedandexplainedtohighlightkeytrendsand patterns.5.1.RequiredAgeDistribution:

TheRequiredAgecolumncategorizesgamesbasedontheminimum recommendedageofthetargetaudience.Thisfeatureensuresthatthe systemfiltersgamesappropriatelyforusersacrossdifferentagegroups. ThedistributionofagecategoriesisillustratedinFigure1,apiechartthat revealstheproportionalrepresentationofagegroupswithinthedataset.

- 44.2%ofgamesaresuitableforusersaged11–15,thelargestagegroup. ThesegamestypicallyincludegenressuchasAdventure,Casual,and Educational,whichareengagingyetappropriateforyoungeraudiences. Thehighproportionreflectsthedemandforfamily-friendlyand educationalgames.
- 32.4%ofgamestargetusersaged16–18.Gamesinthiscategoryare oftenmorecomplexandincludeaction-packedgenreslikeAction,RPG, andStrategy.Thesegamescatertoolderteenagerswhoprefer immersivegameplayexperiences.
- 23.4%ofgamesareintendedforusersaged>18,includingmaturetitles ingenressuchasSurvival,Horror,andSimulation.Thesegamesmay containthemesormechanicsunsuitableforyoungeraudiences.
- Anegligiblepercentage(0%)ofgamesarecategorizedforusersaged

≤10.Thisabsenceindicatesagapinthedatasetforgamestailoredto
veryyoungaudiences,whichcouldbeaddressedbyaugmentingthe datasetwithadditionaltitles.

Page13of50 Figure1:RequiredAgeDistribution
Thechartunderscoresthedataset’semphasisongamesforteenagersand
youngadults.Thesystemusesthisinformationtoimplementage-based
filtering,ensuringusersreceiveage-appropriaterecommendations.
5.2.GameRatingsAnalysis: Ratingsareacriticalmetricinevaluatingthequalityandpopularityof
games.Thedatasetincludesratingsonascalefrom1to10,representing
userfeedbackandreviews.Figure2(boxplot)andFigure3(histogram)
providecomplementaryviewsoftheratingsdistribution. 5.2.1.BoxplotAnalysis:
TheboxplotinFigure2visualizesthespread,centraltendency,and variabilityingameratings:
Themedianratingisapproximately7,indicatingthatmostgamesareof goodquality.
Theinterquartilerange(IQR)spansratingsbetween6and8,which
constitutesthemajorityofthedataset.
Outliersexistatbothextremes.Forinstance,gamesratedbelow4may
representnicheorpoorlyreceivedtitles,whilethoseratedabove9are
exceptionalinqualityanduserappeal. Theseinsightsensurethesystemprioritizeshigh-
ratedgameswhile consideringnichetitlesforuserswithspecificpreferences.

Page14of50 Figure2:RatingsBoxplot 5.2.2.HistogramAnalysis:
ThehistograminFigure3complementstheboxplotbyshowingthe
frequencydistributionofratings:
Ratingsof7and8arethemostcommon,with~300gamesinthese
categories.Thishighlightsthedataset'sfocusonwell-receivedgames, ensuringahigh-
qualityrecommendationpool.
Ratingsbelow5arerare,withonlyafewgamesfallingintothiscategory.
Thesegamesarelikelytobeexcludedordeprioritizedduring recommendationgeneration.

Amodestnumberofgamesachieveaperfectratingof10,representing top-tiertitleshighlyfavoredbyusers. Figure3:GameRatingsDistribution

Thisanalysisdemonstratesthesystem'semphasisonquality,ensuringthat recommendedgamesalignwithuserexpectationsforengagingandenjoyable experiences. 5.3.GenreAnalysis: Genresareadefiningfeatureofgames,capturingtheirthematicand gameplaycharacteristics.Thedatasetincludesuptosevengenresper game,enablingmulti-genreclassifications.Figure4illustratesthetop10 mostfrequentgenres,offeringinsightsintouserpreferencesanddataset diversity.

Actionisthemostdominantgenre,appearinginover500games.This reflectsitsuniversalappeal,spanningvariousagegroupsandgameplay styles.

Indierankssecond,showcasingthegrowingpopularityofindependent gamesthatemphasizecreativityanduniquemechanics.

Adventurefollowsclosely,cateringtouserswhoenjoyexploratoryand narrative-drivenexperiences. OthernotablegenresincludeRPG,Simulation,andStrategy,which appealtousersseekingdepthandcomplexityingameplay.

EarlyAccess,whilelessfrequent,representsgamesindevelopmentthat attractusersinterestedinbetatestingandearly-stageexperiences.

Educationalemergesasanoteworthygenre,targetingusersinterested incombininglearningwithentertainment.Thesegamesofteninclude puzzles,simulations,orstorytellingelementsdesignedtoteachwhile engagingtheplayer.Educationalgamesareprioritizedforuserswho explicitlyselectthisgenreasapreference,emphasizingthesystem'sgoal ofpromotingbothfunandlearning. Figure4:Top10Genres TheinclusionofEducationalgamesunderscoresthesystem'sfocuson promotinglearningthroughgaming,cateringtouserswhoprioritize

intellectualgrowthalongsideentertainment.Thisdiversityallowsthe recommendationsystemtoclusterandfiltergameseffectivelybasedonuserselectedpreferences ,ensuringtailoredsuggestionsthatresonatewith individualtastesandneeds.

5.4.ClusteringInsights: Clusteringplaysapivotalroleintherecommendationsystem,grouping gamesbasedonsemanticsimilarityandsharedattributes.The hierarchicalclusteringdendrograminFigure5visualizesrelationships betweengamesderivedfromtheirtextualdescriptions. 5.4.1.Methodology: TF-IDFVectorization:Thistechniqueconvertsgamedescriptionsinto numericalrepresentations,quantifyingtheimportanceofwordsrelativeto thedataset.

CosineSimilarity:Measuresthesimilaritybetweendescriptions, groupingsemanticallysimilargamesintoclusters. HierarchicalClustering:Constructsatree-likestructure(dendrogram)to representthenestedrelationshipsbetweengames.

5.4.2.Observations: Gameswithinthesameclusteroftensharegenres,themes,orgameplay mechanics.Forinstance,aclustermightincludetitleslike"ActionAdventure"or"Simulation-Strategy." Clusteringreducesredundancybyunifyingvariationsingamenames(e.g., "DeluxeEdition"and"OriginalVersion")underasinglerepresentative entry.

Thedendrogramrevealshigher-levelgroupings,suchasclusters dominatedby"Educational"gamesor"Multiplayer"titles.

Figure5:HierarchicalClusteringDendrogram

Page17of50 Theclusteringanalysisensuresthatrecommendationsarecontextually relevant,offeringusersacuratedlistofgamesthataligncloselywiththeir interests.5.5.KeyInsightsandImplications:

Theinsightsderivedfromthedatasetanalysisareinstrumentalinshaping therecommendationsystem: 1.Age-SpecificFiltering:Theagedistributionensuresthesystem accommodatesdiverseuserdemographics,deliveringage-appropriate content.

2.QualityAssurance:Theprevalenceofgamesrated6–8highlightsthe dataset'sfocusonquality,supportingreliablerecommendations.

3.DiverseGenres:Thewiderangeofgenresallowsthesystemtocaterto variedpreferences,fromcasualplayerstoenthusiastsofnichegenres.

4.SemanticClustering:Theuseofclusteringenhancesthesystem'sability torecommendgamesthatsharethematicorgameplaysimilarities, increasingusersatisfaction.

6.RecommendationSystemDesign:

ThePersonalizedEducationalGameRecommendationSystemisengineered tocombineuserpreferences,collaborativeinsights,andgamemetadatato delivertailoredrecommendations.Thissectiondissectstherecommendation engine’scomponents,detailingitsmethodology,mathematicalfoundations, andpracticalimplementation.Thesystemintegratesmultiplemachine learningtechniques,suchascollaborativefiltering,clustering,and dimensionalityreduction,alongsidecustomscoringmetricstoensureprecise andmeaningfuloutputs. 6.1.AlgorithmOverview:

Therecommendationsystemcombinesmultipleadvancedtechniquesto providepersonalizedandmeaningfulrecommendations.Theseinclude collaborativefiltering,clustering,TF-IDFvectorization,dimensionality reductionthroughSVD,andweightedscoringmechanisms.Thesystem ensuresthateveryrecommendationisbasedonstructuredandefficient computationalmodels.

6.1.1.CollaborativeFiltering(CF):

CollaborativeFiltering(CF)isacornerstoneoftherecommendationengine, enablingittoidentifygamessimilartothoseplayedbytheuser.This techniquereliesontheassumptionthatuserswhoenjoycertaingamesare likelytoenjoyotherswithsimilarcharacteristics. Thesystememploysitem-basedCF,wherethesimilaritybetweengamesis measuredusingcosinesimilarity.Theformulaforcosinesimilarityisas follows: □ □□ □ □ □□

2n1i 2n1in1i =ilarity Cosine Sim i i ii B ABA

Here,AandBrepresentfeaturevectorsfortwogames.Thiscalculation createsasimilaritymatrixthatcapturestherelationshipsbetweenallgamesin thedataset.

Page18of50 Forexample: ConsiderGameAwithfeatures[1,0,1,0].

GameBhasfeatures[0,1,1,1]. 408.0321 11100101 0·1+1·1+0·1+1·0 =ilarity Cosine Sim

2222 2 22 2

Thissimilarityscoreensuresthattherecommendationsaregroundedinthe

relationshipsbetweengamefeatures,enhancingaccuracy. 6.1.2.ClusteringwithK-Means:

Tostreamlinetherecommendationprocess,gamesaregroupedintoclusters

basedontheirfeatures.ThesystemusesK-MeansClustering,which

partitionsdataintokclustersbyminimizingintra-clustervariance. StepsinK-Means:

1.Initialization:kcentroidsarechosenrandomlyinthefeaturespace.

2.Assignment:Eachgameisassignedtothenearestcentroidusing Euclideandistance:

ni iicx1 2 = Distance 3.Update:Thecentroidsarerecalculatedasthemeanofallpointsintheir

cluster. 4.Iteration:Steps2and3arerepeateduntilthecentroidsstabilize.

TheElbowMethoddeterminestheoptimalnumberofclusters(k)byplotting

theSumofSquaredDistances(SSD)againstvariouskvalues.Thepoint

wherethecurveflattensindicatesthebestk,balancingperformanceand computationalefficiency.

6.1.3.TF-IDFVectorization: TheTF-IDF(TermFrequency-

InverseDocumentFrequency)methodconverts text-

basedgamedescriptionsintonumericalrepresentations.Thisensures

thattherecommendationenginecapturesthesemanticessenceofeach

gamewhilereducingtheinfluenceoffrequentlyoccurring,less-informative words. d)·IDF(t)

TF(t,=d) IDF(t, -TF Where: TF(t,d):Measureshowoftentermtappearsindocumentd: s in d

Total term in d Count of t =d)TF(t, IDF(t):Weighsdowntermsappearinginmanydocuments:

t containing Documents ments Total docu log=IDF(t)

Page19of50 Forexample: If“strategy”appearsin2outof10documents:1.70=)102 log(=gy)

IDF(strate Thisensuresthatuniquetermslike"adventure"or"open-world"havegreater

influencethangenerictermslike"game." 6.1.4.SingularValueDecomposition(SVD):

Toreducethedimensionalityofthecombinedfeaturematrix(genresandTFIDF),thesystemapplie sSingularValueDecomposition(SVD).SVDbreaks downthematrixMintothreecomponents:T

U·S·V =M Where: U:Capturesrelationshipsbetweengames.

S:Diagonalmatrixcontainingsingularvalues(featureimportance).

TV:Encodesrelationshipsbetweenfeatures.

Byretainingonlythetopksingularvalues,SVDreducesnoiseandenhances

computationalefficiency.Forexample,retainingthetop50singularvaluesin

amatrixofsize1000×500reducestheeffectivedimensionalitywhile

maintainingkeydatapatterns. 6.1.5.WeightedScoringMechanism:

Thefinalscoreforeachgameiscomputedusingaweightedsumofmultiple factors:

1.SimilarityScore(20%):Captureshowcloselythegamematchesthe user’splayedgames.

2.GenreOverlap(50%):Measuresalignmentbetweentheuser’spreferred

genresandthegame’sgenres: s User Genre Number of enres Matching G Number of =lap

Genre Over 3.Rating(30%):Incorporatesthegame’saveragerating. 0.3·Rating + Overlap

0.5·Genre + rity Score 0.2·Simila =e Final

ScorThisweightedmechanismensuresbalancedandmeaningful recommendations.

6.2.SystemWorkflow: Therecommendationsystem'sworkflowconsistsofasequenceof

interconnectedstepsdesignedtotransformrawdataintopersonalized

gamesuggestionsfortheuser.Belowisabreakdownofeachstageinthe

workflow,asdepictedintheflowchart(Figure6):

Page20of50 Figure6:RecommenderSystemWorkflow

Therecommendationprocessinvolvesfiltering,scoring,andranking

gamesbasedontheuser’sinput.Theworkflowisdesignedtomaximize

relevanceandusersatisfaction. 6.2.1.Step1:UserInput:

Therecommendationprocessbeginswithcollectingcriticalinformationfrom

theuser.Thisstepensuresthesystemcapturestheuser’spreferencesand

personalcontext,formingthebasisforpersonalizedrecommendations.

AgeGroupSelection:Usersspecifywhethertheyareinthe"below15"or

"15andabove"agegroup.Thisstepisessentialforensuringthat recommendationsareage-appropriateandadheretocontent restrictions.Forinstance,usersunder15willnotseegamestaggedwith a15+or18+agerequirement.

PlayedGamesList:Usersprovidethenamesofuptofivegamesthey haveplayed.Thesegamesserveasareferencepointforanalyzingtheir gaminginterestsandidentifyingsimilargamesinthedataset.

FavoriteGenresSelection:Usersselectuptothreepreferredgenres, suchasAction,Adventure,orStrategy.Thesechoicesallowthesystem toprioritizegamesthatalignwiththeuser’stastes.

Thisuser-provideddataservesasthefoundationforsubsequentstepsinthe workflow,allowingthesystemtopersonalizeitsrecommendationseffectively.

6.2.2.Step2:DataFiltering:

Oncetheuserinputsaregathered,thesystemfiltersthedatasettonarrow downthelistofgames.Thisstepeliminatesgamesthatdonotmeetthe user’sbasiccriteriaandensuresonlyrelevantoptionsremain.

AgeFiltering:Gameswitharequired_agehigherthantheuser’s specifiedagegroupareexcludedfromconsideration.Forexample,ifa userbelow15selectsgames,thesystemautomaticallyexcludesgames taggedassuitableforplayersaged15orolder. Genre-BasedPrioritization:Thesystemidentifiesgamesthatshare genreswiththeuser’sfavoritegenres.Forexample,iftheuserselects StrategyandRole-Playing,thesystemprioritizesgamesthatbelongto thesecategories.Thisprioritizationincreasesthelikelihoodof recommendinggamesthatresonatewiththeuser’spreferences.

Filteringsignificantlyreducesthepoolofgamesandensuresonlythose relevanttotheuser’sagegroupandgenresareconsideredforfurther analysis.

6.2.3.Step3:SimilarityComputation:

Topersonalizerecommendationsfurther,thesystemcalculatessimilarity scoresforallgamesinthefiltereddataset.Thissstepleveragesmathematical techniquestoidentifygamesthatsharefeatureswiththeonestheuserhas played. TF-IDFMatrixCreation:ThesystemfirstgeneratesaTermFrequencyInverseDocumentFrequency(TF-IDF)matrixusingthedescriptionsor metadataofthegames.Thismatrixnumericallyrepresentsthe importanceoftermswithinthegamedescriptions,emphasizingunique anddistinguishingfeatures.

CosineSimilarityCalculation:Thecosinesimilarityformulaisapplied tomeasurethesimilaritybetweenvectorsintheTF-IDFmatrix.It calculateshowcloselythedescriptionsoftwogamesalign.Forexample, agamewithhighsimilaritytoauser-selectedgamemayshare keywordslike"multiplayer,""action,"or"strategy."

ScoreAggregation:Thesimilarityscoresbetweeneachuser-selected gameandallothergamesareaveraged.Thisapproachensuresthat thefinalsimilarityscorereflectshowwellagamealignswiththeuser's overallgaminghabitsandpreferences.

Thisstepiscriticalforidentifyinggamesthatnotonlysharegenresbutalso alignwiththethemes,features,anddescriptionsofthegamestheuserhas alreadyenjoyed.

6.2.4.Step4:WeightedScoring:

Torankthegames,thesystemcalculatesaweightedscoreforeachgame basedonmultiplefactors.Thisensuresthatthefinalrecommendations balancesimilarity,userpreferences,andoverallquality.

SimilarityScores:Thesescores,derivedfromthecosinesimilarity calculations,contribute20%tothefinalscore.Thisweightingensures

thatgamessimilartotheonestheuserhasplayedareincludedbutnot overlydominant. GenreOverlap:Genrealignmentwiththeuser’spreferencescontributes 50%tothescore.Thishighweightingreflectstheimportanceof matchinggamestotheuser’sstatedfavoritegenres,ensuringrelevance.

GameRatings:Theratingsofgamesfromthedatasetcontribute30%to thefinalscore.High-ratedgamesaremorelikelytoberecommended, astheyareconsideredbetterqualityandmoreenjoyable.

Theweightedscoringformulaensuresabalancedapproach,combining multipleaspectsofagame’srelevanceandqualityintoasinglescore.This stepiscrucialforrankingthegameseffectively. 6.2.5.Step5:RecommendationOutput: Afterscoring,thesystemselectsthetopfivegameswiththehighestfinal scores.Thesegamesarepresentedtotheuserinanorganizedanduserfriendlyformat.

GameTitle:Eachrecommendationincludesthenameofthegame, allowingtheusertorecognizeiteasily.

AgeRequirement:Theagesuitabilityofthegameisclearlydisplayed, ensuringtheuserknowswhetherthegameisappropriatefortheirage group.

Ratings:Theaverageuserratingisincludedforeachgame,providing insightintoitsqualityandpopularityamongotherplayers.

KeyFeatures:Additionalinformation,suchasthegenresandstandout featuresofthegame,ishighlightedtoexplainwhyitwasrecommended.

·Therecommendationsarepresentedinawaythatemphasizesclarityand usability.Byprovidingdetailedinformationforeachgame,thesystemensures userscanmakeinformeddecisionsaboutwhichgamestoexplorefurther.

6.3.ExplanationofTablesandCalculations:

Thissectionelaboratesonthenumericalcomputationsandinterpretations ofthekeymatricesusedintherecommendationsystem.Eachmatrixis vitalinprocessingdataandderivingmeaningfulrecommendations.Below aretheextendedtablesforTF-IDFMatrix,SVDResults(U,Σ,V^T Matrices),ReconstructedMatrix,andCosineSimilarityMatrix,with additionalrowsandcolumnsforaclearerunderstandingoftheir implications. 6.3.1.TF-IDFMatrix(First15×15Slice): TheTF-IDFmatrixrepresentstheimportanceofdifferentgenres,keywords, orattributesforeachgameinthedataset.Eachrowcorrespondstoagame, andeachcolumnrepresentsafeatureextractedfromthegamedescriptions. Thevaluesinthismatrixreflecttheweightorimportanceofeachfeaturetoa specificgame,calculatedusingtheTermFrequency-InverseDocument Frequencytechnique. Forexample,ifGame1hasaTF-IDFvalueof0.25forthe"Adventure"genre, itindicatesthat"Adventure"isahighlyrelevantcharacteristicofthisgame basedonitsdescription.Conversely,avalueof0for"Puzzle"inthesamerow impliesthat"Puzzle"isirrelevanttothisgame.TheTF-IDFweightshelp

emphasizefeaturesuniquetoagamewhiledownplayinggenericones, ensuringthematrixcapturesthegame'suniqueaspects. Thismatrixplaysafoundationalroleinidentifyingpatternsandrelationships betweengames.Byquantifyingthetextualinformationinthedataset,itallows thesystemtocomputemeaningfulsimilaritiesbetweengames,whichare essentialforgeneratingrecommendations. TheTF-IDFmatrixrepresentstheweightedimportanceoftermsingame descriptions. Game Index StrategyAdventure RPG Multiplayer OpenWorld ActionSimulation Puzzle ShootingSurvivalSandboxHorror Sports Sci-Fi Fantasy Game1 0.180.250.2 0.1 0.150.220.120.00.10.050.080.00.120.070.15 Game2 0.1 0.150.0 0.2 0.050.080.180.00.150.0 0.00.10.00.20.12 Game3 0.0 0.220.18 0.0 0.20.150.0 0.10.150.180.00.080.10.150.1 Game4 0.150.180.1 0.250.150.20.120.10.180.050.10.080.120.10.15 Game5 0.12 0.10.150.150.220.080.180.120.00.120.10.120.080.20.18 6.3.2.SVDResults: SingularValueDecomposition(SVD)isappliedtotheTF-IDFmatrixtoreduce itsdimensionality.Thistechniquebreaksthematrixintothreecomponents:U,

□,andTV,eachservingadistinctpurpose.

TheUmatrixmapsgamestothenewlatentfeatures(principalcomponents)

discoveredduringdimensionalityreduction.Eachrowcorrespondstoagame,

whilethecolumnsrepresentthesecomponents.Thevaluesindicatehow

stronglyagamealignswitheachcomponent. 6.3.2.1.UMatrix(15×15Slice):

Forinstance,ifGame1hasavalueof0.065inthefirstcolumnofthe

Umatrix,itmeansthatthefirstprincipalcomponentsignificantlyinfluences

thisgame'srepresentation.Theseprincipalcomponentsareabstractfeatures

thatcombineoriginalattributessuchasgenresorthemes.Forexample,a

singlecomponentmightencapsulateacombinationof"Adventure"and "Strategy."

TheUmatrixreducesthecomputationalcomplexityofsimilarity

measurements.Insteadofcomparinggamesacrosshundredsoffeatures,the

systemnowworkswithafewhigh-levelcomponents,enablingfasterand

moreefficientcalculations.

Page24of50 RowIndexComponent 1 1 Component 2 Component 3 Component 4

Component 5 Component 6 Component 7 Component 8 Component 9 Component 10

Game1 0.065 -0.04 0.03 -0.0280.056 0.07 -0.0450.031 0.051 0.062 Game2 0.045

0.053-0.0350.022-0.0580.012 0.062-0.0380.044-0.046 Game3 0.03 0.02 0.065 -0.04 0.035

0.025-0.046 0.06 0.03 -0.032 Game4 -0.0280.022 0.02 0.063-0.0520.048-0.0430.056

0.062 -0.04 Game5 0.056-0.058-0.0450.035 0.03 -0.0340.058 0.051 -0.04 0.065

6.3.2.2.ΣMatrix(SingularValues):

The□matrixcontainsthesingularvalues,whichrepresenttheimportance

ofeachprincipalcomponent.Largervaluesindicatecomponentsthatcapture

morevarianceorinformationfromtheoriginaldata.Forinstance,thefirst

componentmighthaveasingularvalueof28.02,signifyingitsdominancein

representingthedataset.Incontrast,componentswithsmallervalues,such

as12.40,contributelessandareoftendisregardedduringdimensionality reduction.

Byfocusingoncomponentswithhighsingularvalues,thesystemretainsthe mostmeaningfulpatternsinthedatawhilediscardingnoiseandredundant information.Thisensuresthattherecommendationsystemoperates efficientlywithoutsacrificingaccuracy. Component Value 1 28.02 2 21.45 3 19.3 4 15.8 5 12.4 6.3.2.3.VTransposed(V^T)Matrix:

TheTVmatrixrelatestheoriginalfeatures(genresandkeywords)tothe principalcomponents.Eachrowrepresentsafeature,andeachcolumn representsaprincipalcomponent.Thevaluesinthismatrixindicatethe contributionofeachfeaturetoaspecificcomponent.

Forexample,ifthe"Strategy"genrehasavalueof0.62inthefirstcolumn,it meansthisgenreheavilyinfluencesthefirstprincipalcomponent.Negative valuesindicateaninverserelationship,meaningfeatureswithnegative weightsdetractfromthecomponent'srepresentation.UnderstandingtheTV matrixallowsustointerpretthelatentfeaturesandidentifywhich combinationsofgenresorattributesdrivesimilaritiesbetweengames.

Page25of50 FeatureComponent 1 Component 2 Component 3 Component 4 Component 5 Strategy 0.62 0.45 -0.12 0.34 0.41 Adventure 0.52 0.21 0.31 -0.48 -0.32 RPG 0.33 0.55 0.18 0.25 -0.4 Multiplayer 0.45 -0.3 0.48 0.22 0.5 Open-World 0.29 0.38 -0.25 -0.39 0.48 6.3.3.ReconstructedMatrix(15×15Slice):

ThereconstructedmatrixapproximatestheoriginalTF-IDFmatrixafter dimensionalityreduction.BymultiplyingtheU,□,andTVmatrices,we obtainacompressedversionoftheoriginaldata.Thisreconstructioncaptures themostsignificantrelationshipsbetweengamesandfeatureswhile discardinglessrelevantdetails.

Forexample,ifthereconstructedvaluefor"Strategy"inGame1is0.18,it closelyalignswiththeoriginalTF-IDFvalueof0.20,indicatingthatthe dimensionalityreductionpreservedthecriticalinformationforthisfeature.

Minordiscrepanciesbetweentheoriginalandreconstructedvaluesare expectedandrepresentthetrade-offbetweendatacompressionand informationretention. Thereconstructedmatrixvalidatestheeffectivenessofthedimensionality reductionprocess.Itensuresthatthecriticalaspectsofthedataare preserved,enablingthesystemtogeneratereliableandmeaningful recommendations. Game IndexStrategyAdventureRPGMultiplayerOpenWorldActionSimulationPuzzleShootingSurvivalSandboxHorrorSportsSci-FiFantasy Game1 0.180.260.220.10.120.230.140.10.110.070.080.10.080.120.13 Game2 0.120.180.10.20.10.180.150.10.10.120.120.080.090.10.14 Game3 0.080.220.20.10.250.20.080.120.110.10.120.150.110.090.12

6.3.4.CosineSimilarityMatrix(15×15Slice):

Thecosinesimilaritymatrixquantifiesthesimilaritybetweengamesbasedon theirfeaturerepresentations.Eachrowandcolumncorrespondstoagame, andthevaluesrangefrom-1to1.Avalueof1indicatesperfectsimilarity, whileavalueof0signifiesnosimilarity.Negativevalues,thoughrareinthis context,wouldindicatedissimilarity.

Page26of50 Forinstance,Game1hasasimilarityscoreof1.00withitself,asexpected. However,itmightalsohaveascoreof0.75withGame3,indicatingthatthese gamessharemanycommonfeatures,suchasoverlappinggenresorsimilar descriptions.Incontrast,ascoreof0.35withGame2suggestsfewershared features. Thismatrixiscriticalfortherecommendationengine.Bycomparingauser's previouslyplayedgamestotherestofthedataset,thesystemidentifiesthe mostsimilargamesandrecommendsthem.Thecosinesimilaritymatrix ensuresthatrecommendationsarenotonlyrelevantbutalsopersonalizedto theuser'spreferences. GameIndexGame1 Game2 Game3 Game4 Game5 Game1 1.0 0.48 0.65 0.53 0.59 Game2 0.48 1.0 0.45 0.52 0.62 Game3 0.65 0.45 1.0 0.47 0.55 Game4 0.53 0.52 0.47 1.0 0.61 Game5 0.59 0.62 0.55 0.61 1.0 7.Implementation:

TheimplementationofthePersonalizedEducationalGameRecommendation Systemwasstructuredintothreeoverarchingphases:datapreparation, systemdesign,andinterfacedeployment.Eachphaseinvolvedmeticulous attentiontodetailtoensureprecision,scalability,andusabilityofthefinal recommendationengine. 7.1.ToolsandLibrariesUsed:ComprehensiveAnalysis: Tool/LibraryPurposeintheSystem WhyThisToolWasChosen Python Coreprogramming language. Pythonisversatile,widelyusedinmachine learning,andhasrichlibrariesfordata analysis,textprocessing,andapplication development.Itenablesseamlessintegration ofmultiplefunctionalities,frompreprocessing datasetstodeployingwebapplications. pandas Datamanipulation: LoadingCSVs,cleaning missingvalues,and groupingdata. pandassimplifiesoperationslikefiltering, merging,grouping,andpivoting,essentialfor preparingthedatasetforclusteringand recommendation.Ithandlesmissingand inconsistentdataefficiently,whichiscrucial whenprocessinglargegamedatasets. numpy High-performance matrixoperations. numpyacceleratesnumericalcomputations, whicharecriticalforbuildingsimilarity matrices,performingSVD,andnormalizing vectorsintheTF-IDFprocess.Itslightweight natureensuresscalabilityforlargedatasets.

Page27of50 scikit-learn Machinelearning algorithmslikeTF-IDF, DBSCAN,andSVD. scikit-learnoffersproduction-grade implementationsofclusteringand dimensionalityreductionalgorithms.These algorithmsarecentraltogroupinggames, extractingmeaningfulfeaturesfromtext,and reducingcomputationoverheadthroughSVD. Streamlit Buildingtheinteractive userinterface. Streamlitenablesrapidprototypingofweb applications.Itabstractsthecomplexityof front-enddevelopmentwhileofferingdynamic widgetsforuserinput,suchasmultiselect dropdownsandsliders.Itallowedusto quicklycreateapolished,interactiveinterface forusers. re(Regular Expressions) Textcleaningand preprocessing: Removingspecial charactersand normalizingtextfields.Regularexpressionsareindispensablefor stringoperations.Theyensureuniformityin gamenamesanddescriptions,eliminating inconsistenciesthatcouldaffectclusteringor similaritycalculations. matplotliband seaborn Visualizationofdataset trendsandclustering results. Datavisualizationtoolslikematplotliband seabornwereessentialforexploringpatterns, suchasgenredistribution,ratinghistograms, andsimilarityheatmaps.Theyalsoprovided insightsintoclusteringandage-group distributions,improvingtheinterpretabilityof thedatapreprocessingphase. math Mathematical operationsforTF-IDF andinversedocument frequencycalculations. Themathematicalunderpinningsof recommendationsystems,suchascalculating IDFscoresorcosinesimilarity,required preciselogarithmicandsquareroot calculations,whichthemathlibraryfacilitates withefficiencyandreliability. TF-IDFVectorizeTransformingtextual descriptionsinto numerical representations. TheTF-IDFVectorizerfromscikit-learn convertstextintovectorsthatquantifyword importance.Thiswascriticalforcomparing gamedescriptionssemanticallyand clusteringsimilargames. DBSCAN (Density-Based Spatial Clustering)Clusteringgamenames tounifyvariationslike “DeluxeEdition.” DBSCAN’sabilitytoclusterbasedondensity ratherthanfixedclusterswasidealfor groupinggames.Ithandlesnoiseeffectively, allowingittodistinguishbetweenactualgame clustersandunrelatedentries. SVD(Singular Value Decomposition) Dimensionality reductionofTF-IDF vectorsfor computational efficiency. SVDreducesthedimensionalityoftext features,retainingthemostrelevant componentswhilediscardingnoise.Thisstep wasvitalforcreatingacosinesimilarity matrixwithoutoverloadingmemoryor computationalresources.

7.2.ImplementationProcess:Step-by-StepAnalysis: Theimplementationprocessinvolvedmultiplestages,eachbuildingupon thepreviousto createarobustandreliablerecommendationsystem. Belowisadetailedanalysisofeachstep,providingacomprehensive understandingoftheworkflow. 7.2.1.Step1:DataPreparationandCleaning: Preparingthedatasetwasafoundationaltask.Therawdatacontained severalinconsistencies,redundantcolumns,andmissingvalues.Addressing theseissueswascriticaltoensurethequalityofdownstreamalgorithms. 1.ColumnSelection:

Certaincolumns,suchasis_free,header_image,andwebsite,were deemedirrelevanttotherecommendationlogicanddropped.

Keycolumns,includingname,short_description,genres,andratings, wereretainedastheyheldessentialinformationforclustering, similaritycomputation,anduser-specificrecommendations. 2.HandlingMissingData:

Forcriticalcolumnslikeratings,missingvalueswerereplacedwith defaultplaceholdersorcalculatedaverages.

Forlesscriticalcolumns(e.g.,optionalgenrefields),missingvalues wereignoredduringclusteringandfeatureengineering.

3.TextNormalization:Textfields,particularlygamenamesanddescriptions, werestandardizedby: Lowercasingalltexttoensurecase-insensitivematching.

Removingspecialcharacters,numbers,andextraspacesusingregular expressions.

Tokenizingandstemmingwordsindescriptionstoreduceredundancy andfocusontherootmeaningsofterms.

Example:Thegamename"Minecraft:DeluxeEdition"wastransformedinto "minecraftdeluxeedition."Thisensuredconsistencyinhowgamenames wererepresentedandeliminatedvariationsthatcouldmisleadclustering algorithms.

7.2.2.Step2:ClusteringGameNames:

Duplicateandversionedgamenamesposedasignificantchallenge.For example,"FIFA20,""FIFA2020Deluxe,"and"FIFA20UltimateEdition" representedthesamebasegamebutwerelistedseparately.Tounifysuch entries: 1.TF-IDFVectorization: Gamenameswereconvertedintonumericalvectors,capturingtheir uniquetextualfeatures. TF-IDFweightedtermsbasedontheirfrequencywithinagamename andtheirrarityacrossallnames,enablingeffectivedifferentiation betweenunrelatednames.

2.DBSCANClustering:

Usingcosinesimilarityasthedistancemetric,DBSCANgrouped similarnamesintoclusters. Noisehandlingwasakeystrengthof DBSCAN,ensuringunrelated entrieswereexcludedfromclusters. Thedensity-basednatureofDBSCANeliminatedtheneedtopredefine thenumberofclusters,allowingdynamicadjustmentsbasedonthe data.

3.RepresentativeNameSelection:Withineachcluster,theshortestname waschosenastherepresentative.Forinstance,"FIFA20"wasselectedto representallvariationsofthegame. WhyDBSCAN?

- DBSCANeffectivelyhandlesdatawithnoiseandoutliers,acommon occurrenceinlargedatasets.
- Itgroupsentriesbasedondensity,makingitidealforclusteringtextual datawithvariablesimilaritythresholds. 7.2.3.Step3:FeatureEngineering:

Featureengineeringinvolvedcreatingnumericalrepresentationsofthe datasettoenablemachinelearningalgorithmstounderstandthedata.

1.GenreEncoding:Gamegenreswereone-hotencodedintoabinarymatrix. Forexample:

- Agamewithgenres"Action"and"Adventure"wasrepresentedas[1,1, 0,0,...].
- Agamewithonly"Action"was[1,0,0,0,...].

Thisrepresentationallowedthesystemtomeasuregenreoverlapbetween games.

2.TextVectorization: Descriptionswereconvertedintonumericalvectorsusing TF-IDF.This approachhighlightedtheimportanceofuniquetermsinagame's descriptionrelativetoothers.

- Stopwordssuchas"game,""play,"and"new"wereremovedtofocus onmeaningfulterms.

7.2.4.Step4:DimensionalityReductionUsingSVD: TheTF-IDFmatrixoftencontainedthousandsofdimensionsduetothe extensivevocabulary.Tomakecomputationsefficient: 1.SingularValueDecomposition(SVD):

- Thematrixwasreducedto30components,capturingthemost significantfeatureswhilediscardingnoise.
- Thisstepdramaticallyreducedthecomputationalcomplexityof similaritycalculations.

2.MathematicalInsight: TheTF-IDFvalueforeachtermwascalculatedas:IDF×TF=IDF-TF

Where:

Page30of50ent s in docum Total term document of term in Frequency =TF with term

Documents ments Total docu log=DF □□□ □□□

SVDdecomposedthematrixintothreecomponents:U,□,andTV,

reducingthedimensionalitywhileretainingkeypatterns.

7.2.5.Step5:BuildingtheRecommendationEngine:

Therecommendationenginecombinedmultiplefeaturestocomputesimilarity andrankgames.

1.SimilarityMatrix:UsingthereducedmatrixfromSVD,acosinesimilarity

matrixwascalculated.Thismatrixquantifiedhowcloselyeachgame resembledothers.

2.WeightedScoring:Aweightedformulacombinedthreekeyfactors:

□20%:Similaritytogamestheuserhadplayed. □50%:Genreoverlapwiththeuser'spreferences.

□30%:Averageratingofthegame.

Thisapproachensuredtherecommendationsbalancedrelevance,user preferences,andquality.

7.2.6.Step6:UserInterface: Thefinalphasewascreatinganintuitiveinterfaceforusers.

1.StreamlitIntegration:Theinterfacealloweduserstoinput: □Theiragegroup.

□Alistofupto5gamestheyhadplayed. □Upto3preferredgenres.

Usersreceiveddetailedrecommendationswithinsightsintowhyeachgame wassuggested.

2.DetailedExplanations:Foreachrecommendedgame,theinterface displayed:

□Similarityscorewiththeuser'sinput. □Genreoverlappercentage.

□Gameratingandrequiredage.

□Anyadditionalmetadata,suchassupportedlanguagesandcategories.

UserInteraction:Theinterfaceoffereddynamicfeedback,suchaswarnings

iftheuserselectedfewergamesorgenresthanrequired.Recommendations

werepresentedwithvisualaids,includinggamecoverimageswhere

available.7.3.CodeSnippetsandDetailedExplanations: 7.3.1.TF-IDFVectorization:

Thefirstcriticalstepincreatingtherecommendationsysteminvolved

convertinggamedescriptionsintonumericalvectorsusingTermFrequencyInverseDocumentFr equency(TF-IDF).TF-IDFhighlightstheimportanceof

wordsinadocumentrelativetotheentiredataset,ensuringthatcommon
words,whicharelessinformative,aredown-weighted.Forthistask,weused
theTfidfVectorizerfromthesklearn.feature_extraction.textlibrary.
Theprocessbeginsbyfittingthevectorizeronthe"merged_description"
columnofthedataset.Thiscolumncontainsconsolidateddescriptionsofeach
game.Thevectorizertokenizesthetext,removespredefinedstopwords(e.g.,
“game”or“play”),andassignsweightstoeachtermbasedonitsfrequencyin
aspecificgame’sdescriptionrelativetotheoveralldataset.Forexample,a
gamedescribedas“action-packedmultiplayershooter”mightgeneratea
vectorwheretermslike“action”and“multiplayer”areweightedhigherifthey
occurlessfrequentlyacrossthedatasetbutarepivotaltothisparticulargame.
Thistransformationisvitalasitcapturesthesemanticessenceofeach
game'sdescriptioninanumericalformat,enablingmathematical
computationsandcomparisons. desc_corpus=df_grouped["merged_description"].values
raw_vocab=set() fordocindesc_corpus: words=doc.split() forwinwords: raw_vocab.add(w)
raw_vocab_list=sorted(list(raw_vocab)) custom_sw={
"game","games","play","played","playing","world","players","new",
"open","will","make","makes","made","thing","things","like"
}filtered_vocab_list=[vforvinraw_vocab_listifvnotincustom_sw]
defterm_frequency(doc,vocab): counts=[0]*len(vocab) words=doc.split() forwinwords:
ifwinvocab: idx=vocab.index(w) counts[idx]+=1 returncounts tf_matrix=[]
fordocindesc_corpus: tf_matrix.append(term_frequency(doc,filtered_vocab_list))
tf_matrix=np.array(tf_matrix,dtype=float) doc_freq=np.count_nonzero(tf_matrix,axis=0)
N_docs=len(desc_corpus) final_vocab=[] valid_idx=[]

fori,winenumerate(filtered_vocab_list): ratio=doc_freq[i]/N_docs ifratio<=0.6:

final_vocab.append(w) valid_idx.append(i) tf_matrix_filtered=tf_matrix[:,valid_idx]
df_filtered=doc_freq[valid_idx] idf_vals=[] fordf_valindf_filtered:
idf_vals.append(math.log((N_docs+1)/(df_val+1))+1) idf_vals=np.array(idf_vals)
tf_idf_matrix=tf_matrix_filtered*idf_vals defrowwise_norm(mat): out=[] forrowinmat:
length=np.sqrt(np.sum(row**2)) iflength!=0: out.append(row/length) else:out.append(row)
returnnp.array(out) tf_idf_matrix=rowwise_norm(tf_idf_matrix)
Thisstepensuresthatthetextualdataisnumericallyrepresented,allowingfor
deeperanalysisinlaterstages. 7.3.2.SVDforDimensionalityReduction: OncetheTF-
IDFmatrixisgenerated,ittypicallyspansthousandsof
dimensionsduetothelargevocabularyderivedfromgamedescriptions.To
makecomputationsmoreefficient,SingularValueDecomposition(SVD)is
applied.SVDisamathematicaltechniquethatdecomposesthehighdimensionalmatrixintothreec
omponents:U,□,andTV.These
componentscapturetheessentialpatternsinthedata,discardingnoiseand redundancies.
Thereducedcomponentsarethenusedtoreconstructalower-dimensional
matrix.Byretainingonlythetop30components,weensurethatthemost
significantfeaturesarepreservedwhilesignificantlyreducingcomputational
overhead.Forinstance,gameswithsimilarthemesordescriptionswillcluster
togetherinthisreducedspace,eveniftheyuseslightlydifferentwording.
defsvd_decomposition(matrix,k=None): A=np.array(matrix,dtype=float) ATA=np.dot(A.T,A)
AAT=np.dot(A,A.T)

Page33of50 eigvals_ATA,V=np.linalg.eigh(ATA) eigvals_AAT,U=np.linalg.eigh(AAT)
idx_v=np.argsort(eigvals_ATA)[::-1] idx_u=np.argsort(eigvals_AAT)[::-1] V=V[:,idx_v]
U=U[:,idx_u] eigvals=eigvals_ATA[idx_v] ifkisnotNone: U=U[:,:k] V=V[:,:k]
eigvals=eigvals[:k] sigma=np.sqrt(eigvals) sigma_matrix=np.diag(sigma)
returnU,sigma_matrix,V.T K=30 U,S,Vt=svd_decomposition(combined_features,k=K)
reconstructed=np.dot(np.dot(U,S),Vt)

Thisdimensionalityreductionstepnotonlyimprovestheefficiencyofthe systembutalsoenhancestheaccuracyofsimilaritycomputationsbyfocusing onmeaningfulfeatures. 7.3.3.CosineSimilarityforGameMatching:

Torecommendgamessimilartotheonesauserhasplayed,wecomputethe cosinesimilaritybetweengames.Cosinesimilaritymeasurestheangle betweentwovectorsinahigh-dimensionalspace.Asmallerangle(closerto zero)indicateshighersimilarity.Forinstance,twoactiongameswithsimilar descriptionsandoverlappinggenreswillhavevectorspointinginnearlythe samedirection,resultinginahighsimilarityscore.

Thenormalizedfeaturevectors(fromthereducedmatrix)areusedto computeacosinesimilaritymatrix.Thismatrixrepresentspairwisesimilarity scoresforallgames,formingthebackboneoftherecommendationengine.

```
defcosine_similarity_matrix(features):
norms=np.sqrt(np.sum(features**2,axis=1,keepdims=True))
normed=features/(norms+1e-8) sim=np.dot(normed,normed.T) returnsim
item_sim_matrix=cosine_similarity_matrix(reconstructed)
```

Byleveragingcosinesimilarity,wequantifyhowcloselyrelatedanytwo gamesarebasedontheircombinedfeatures,ensuringthatrecommendations alignwithuserpreferences.

7.3.4.RecommendationScoringMechanism:

Thefinalstepintherecommendationprocessinvolvesaggregatingmultiple metrics—similaritytoplayedgames,genreoverlap,andratings—intoasingle weightedscore.Eachmetricisassignedaweightbasedonitsimportance. Similaritycontributes20%,genreoverlapaccountsfor50%,andratingsadd 30%tothefinalscore.Thisweightingstrategyensuresthatthe recommendationsarenotonlysimilartotheuser’spastchoicesbutalsoalign withtheirgenrepreferencesandmaintainahighqualitystandard.

Forinstance,ifauserhasplayedmultipleactiongames,thesystemwill

prioritizeactiontitleswithhighratingsandastrongoverlapingenres.The

recommendationfunctioncomputesthesescoresdynamically,ranksthe

games,andreturnsthetoprecommendations. defrecommend_games(user_age,

#"below_15"or"above_15" user_games,#listoforiginalnames

user_genres,#listoffavoredgenres top_n=5):ifuser_age=="below_15":

valid_df=df_grouped[df_grouped["required_age"]<=15].copy()

else:valid_df=df_grouped.copy() valid_idx=valid_df.index.tolist() defclean(gm):

returnre.sub(r"[^a-z0-9\s]","",gm.lower()).strip()

user_cleaned=[clean(gm)forgminuser_games] rep_indices=[] forucinuser_cleaned:

ifucinname_to_rep: cluster_rep=name_to_rep[uc]

row_match=valid_df[valid_df["unified_name"]==cluster_rep] iflen(row_match)>0:

ridx=row_match.index[0] rep_indices.append(ridx)

else:row_match=valid_df[valid_df["unified_name"]==uc] iflen(row_match)>0:

ridx=row_match.index[0] rep_indices.append(ridx) iflen(rep_indices)==0:

valid_df["similarity_score"]=valid_df["ratings"].astype(float)

valid_df.sort_values("similarity_score",ascending=False, inplace=True)

returnvalid_df.head(top_n) avg_scores=[] foridxinvalid_idx: local_sims=[]

forr_idxinrep_indices: local_sims.append(item_sim_matrix[idx,r_idx])

avg_scores.append(np.mean(local_sims))

valid_df["similarity_score"]=avg_scores

defmeasure_overlap(row_gs,user_gs): row_set=set(row_gs) user_set=set(user_gs)

overlap_count=len(row_set.intersection(user_set)) iflen(user_gs)==0: return0

returnoverlap_count/len(user_gs) overlap_scores=[] fori,rowinvalid_df.iterrows():

overlap_scores.append(measure_overlap(row["merged_genres"], user_genres))

valid_df["genre_overlap"]=overlap_scores w_sim=0.2 w_genre=0.5 w_rat=0.3

final_score=[] fori,rowinvalid_df.iterrows(): score=(w_sim*row["similarity_score"]+

w_genre*row["genre_overlap"]+ w_rat*row["ratings"])final_score.append(score)

valid_df["combined_score"]=final_score

valid_df.sort_values("combined_score",ascending=False,inplace=True)

top_recs=valid_df.head(top_n).copy() sim_details=[] fori,rowintop_recs.iterrows(): txt=""

forr_idxinrep_indices: rep_name=df_grouped.loc[r_idx,"unified_name"]

val_sim=item_sim_matrix[i,r_idx] txt+=f"Similarto{rep_name}=>{val_sim:.3f}\n"

sim_details.append(txt.strip()) top_recs["similarity_details"]=sim_details returntop_recs

Thisscoringmechanismensuresabalancedrecommendationsystem,

offeringusersdiverseyethighlyrelevantgamesuggestions.

7.4.FlowchartsforImplementationProcess:

Thissubsectionexplainsthedetailedflowcharts(Figures7to12)usedto

representthesystem'soperation.Eachflowchartillustratescritical

componentsoftheimplementationprocess,breakingdowntheworkflow

intomanageableandcomprehensiblesteps. 7.4.1.DatasetPreprocessingFlowchart:

Thefirstflowchartrepresentsthedatasetpreprocessingphase,essentialfor

ensuringtherawdataiscleanandreadyforfurtherprocessing.Theprocess

beginsbyloadingthedataset,whichincludescolumnslikegamenames,

genres,ratings,anddescriptions.Irrelevantcolumnssuchasis_freeand

header_imagearedroppedtofocusonlyonusefulinformation.Missing

valuesincriticalfieldslikeratingsarefilledusingaveragesorplaceholders, whilenon-

criticalfieldswithmissingvaluesareignored.Textfields,including

gamenamesanddescriptions,arenormalizedbyconvertingthemto

lowercase,removingspecialcharacters,andtokenizingthemfor

standardization.Genresareconvertedintoabinarymatrixusingone-hot

encoding,ensuringthattheyaremachine-readable.Theoutputisacleanand

structureddatasetthatisreadyforclusteringandfeatureengineering.

Figure7:DatasetPreprocessing 7.4.2.ClusteringGameNamesFlowchart:

Thisflowchartoutlinestheclusteringprocesstounifysimilargamenamesand reduceredundancy.TheprocessstartsbyapplyingTF-IDFvectorizationto gamenames,convertingthemintonumericalvectorsbasedontheirtextual features.DBSCANclusteringisthenusedtogroupsimilarnamesbasedon theirdensity,usingcosinesimilarityasthemetric.Foreachcluster,a representativename—typicallytheshortestname—isselectedtostandardize gameversions.Theseunifiednamesreplaceoriginalgamenamesinthe dataset,creatingaconsistentdatasetwherevariationslike“FIFA20”and “FIFA2020Deluxe”aregroupedunderasinglename.Thefinaloutputisa datasetwithclusteredgamenames,enablingbetterfeatureengineeringand recommendationaccuracy.

Page37of50 Figure8:ClusteringGameNames 7.4.3.FeatureEngineeringFlowchart: Thefeatureengineeringflowchartdetailsthecreationofmeaningfulfeatures fromthedataset.Theprocessbeginsbymergingallgenresassociatedwith eachgame.Thesegenresarethenencodedintoabinarymatrixusingonehotencoding.Simultane ously,gamedescriptionsarecleanedandnormalized toremoveunnecessarynoise.TF-IDFvectorizationisappliedtothese descriptions,generatingnumericalvectorsthathighlighttheimportanceof specifictermsindescribingeachgame.ThegenrematrixandTF-IDF featuresarethencombinedintoaunifiedfeatureset,capturingboth categoricalandtextualinformation.Theresultingfeaturematrixisusedfor dimensionalityreductionandsimilaritycomputations,ensuringefficientand accuraterecommendations.Figure9:FeatureEngineering

Page38of50 7.4.4.RecommendationFunctionFlowchart: Therecommendationfunctionflowchartvisualizesthelogicusedtogenerate personalizedrecommendations.Theprocessstartswithuserinputs,including agegroup,favoritegenres,andpreviouslyplayedgames.Thefirststepfilters theavailablegamesbasedontheuser’sagegrouptoexcludeage-restricted content.Next,thesystemmatchestheuser’splayedgamestotheir respectiveclusterstoidentifyrelevantdatapoints.Usingcosinesimilarity, similarityscoresarecalculatedbetweentheuser’splayedgamesandthe remaininggames.Genreoverlapisalsocomputedtoassesshowwellagame alignswiththeuser’spreferences.Thesemetricsarecombinedintoa weightedscoringformula,assigning20%weighttosimilarity,50%togenre overlap,and30%toratings.Finally,thetop-rankedgamesareoutputtedas recommendations,accompaniedbydetailedreasoning. Figure10:RecommendationFunction

7.4.5.StreamlitUserInterfaceFlowchart:

Thisflowchartexplainstheinteractionbetweentheuserandthe recommendationsystemthroughaStreamlit-basedinterface.Theuserbegins bylaunchingtheapplicationandselectingtheiragegroup.Theythenchoose uptofivegamestheyhaveplayedanduptothreefavoritegenres.Oncethe userclicksthe“GetRecommendations”button,thesystemvalidatesthe inputstoensureallrequirementsaremet.Thevalidatedinputsarepassedto therecommendationfunction,whichgeneratespersonalizedgame recommendations.Theserecommendationsaredisplayedtotheuser,

completewithexplanationsforeachsuggestion,makingtheinterface interactiveanduserfriendly. Figure11:StreamlitUserInterface

7.4.6.RecommendationEngineLogicFlowchart:

Therecommendationenginelogicflowchartprovidesadeeperviewofthe system'sbackendoperations.Theenginebeginsbyreceivinguserinputs suchasagegroup,playedgames,andpreferredgenres.Itfiltersthegames basedonagerestrictionstomatchtheuser’sprofile.Next,theuser’splayed gamesaremappedtotheirrespectiveclustersforeffectivecomparison.

Clusterrepresentativesarevalidatedtoensureaccuracy.Cosinesimilarity scoresarecalculatedbetweentheuser’splayedgamesandothergames, whilegenreoverlapismeasuredtoevaluatealignmentwiththeuser’s preferences.Thesemetricsarecombinedintoafinalweightedscore,which ranksgamesbasedontheirrelevancetotheuser.Thesystemsorts recommendationsbytheirfinalscoresandoutputsthetopgamesas recommendations,ensuringpersonalizedandaccuratesuggestions.

Page40of50 Figure12:RecommendationEngineLogic 8.TestingandResults: 8.1.Testing: Testingisacriticalphasetoensurethereliability,accuracy,andusability oftheGameRecommenderSystem.Thegoalwastoevaluatethesystem undervariousscenarios,coveringbothfunctionalandnon-functional aspects.Thesystemunderwentextensivetestingacrossmultiple dimensions,includinginputvalidation,algorithmaccuracy,usability,and performance.
8.1.1.TestingObjectives: Theprimaryobjectivewastovalidatethattherecommendersystem consistentlygeneratesaccurate,relevant,anduser-specificrecommendations.
Thesystemwasexpectedto:

Page41of50 Handleedgecases,suchasinvalidorincompleteuserinputs.
Filtergamesbasedonagerestrictionsaccurately.
Providerecommendationsthatreflecttheuser’sgamepreferencesand genreselections.
Deliverclearandmeaningfulexplanationsfortherecommended games.
Performefficientlyundervaryingdatasetsizes. 8.1.2.TestMethodology:
Thetestingwasdividedintothreemainstages:
1.UnitTesting:Focusedonensuringindividualcomponentslikeclustering, collaborativefiltering,andcosinesimilaritycalculationswereworkingas intended.
2.IntegrationTesting:Verifiedthatthebackendalgorithmsintegrated smoothlywiththefrontendStreamlitapplicationtoprovidereal-time results.

3.UserTesting:Conductedwith20participantswhoprovidedinsightsinto usability,recommendationrelevance,andthequalityofexplanations.

8.1.3.TestScenariosandResults:

Thesystemwasevaluatedusingvarioustestcases,asoutlinedbelow: TestScenario ExpectedOutcome Result UserInputValidationPreventincompleteinputs(e.g.,lessthan5 gamesselected)anddisplaywarnings. Passed AgeRestriction Filtering Displaygamesonlysuitablefortheselected agegroup(below_15orabove_15). Passed Recommendation Accuracy Recommendgamessimilartotheuser’s selectedgamesandfavoritegenres. Passed ExplanationClarityProvidedetailedsimilarityandgenreoverlap explanationsforeachrecommendation. Passed Performanceunder Load Generaterecommendationswithin3seconds fordatasetsofupto10,000games. Passed(1.8s avg.) InterfaceUsabilityEnsuredropdownmenus,errormessages, andbuttonsareintuitivetouse. Passed

Theresultsshowthesystemperformedreliablyacrossalltestscenarios, meetingfunctionalandusabilityexpectations. 8.2.ResultsRepresentation: Thesystem’sresultswereevaluatedthroughexampleusecasesto demonstrateitsfunctionalityandexplainability.Belowisanexample scenario: InputDetails AgeGroup:Above15 GamesPlayed:Brain/Out,AgeofHistoryII,Fez,Inside,Teardown FavoriteGenres:Action,Adventure,Simulation RecommendationsOutput

Basedontheinput,thesystemprovidedthefollowingrankedlistof recommendedgames: GameNameRequired Age Rating Similarity Score Genre Overlap Combined Score Assassin’s CreedIV: BlackFlag 18 10.0 0.284 0.667 3.390 Tricolour Lovestory 18 10.0 0.352 0.333 3.237 Bioshock Infinite 18 10.0 0.305 0.333 3.228 Half-Life 18 10.0 0.298 0.333 3.226 EldenRing 18 10.0 0.210 0.333 3.209

Therecommendationsalignedwellwiththeuser’sinput.Gameslike "Assassin’sCreedIV:BlackFlag"and"Half-Life"scoredhighduetotheir overlapwithgenreslikeActionandAdventure.Theirsimilarityscores,

derivedfromthedescription-basedTF-IDFanalysis,wereconsistently

above0.2,whichindicatesastrongalignmentwiththeuser’sselected games.

SimilarityScore:Reflectshowcloselytherecommendedgame

descriptionsmatchthegamesselectedbytheuser.Forexample,

"BioshockInfinite"scored0.305duetoitsnarrative-drivenstylesimilar to"Inside."

GenreOverlap:Measuresthepercentageofgenresthatmatchthe

user’sfavorites."Assassin’sCreedIV:BlackFlag"achievedthe

highestoverlap(0.667)becauseitalignswithallthreegenresselected.

CombinedScore:Aweightedsumofthesimilarityscore(20%),genre

overlap(50%),andgamerating(30%).Thisscoredeterminesthefinal ranking.

8.3.VisualRepresentationofResults: 1.InputInterface:Theuser-

friendlyinterfaceallowsuserstoselectage group,playedgames,andgenresseamlessly:

Figure13:InputInterface

2.RecommendationOutput:Recommendationsaredisplayedwith

detailslikerequiredage,ratings,similarityscores,andexplanations:

Figure14:RecommendationOutput 8.4.InsightsandFutureEnhancements:

8.4.1.Insights: TheGameRecommenderSystemdemonstratedastrongabilitytodeliver

personalized,relevant,andexplainablerecommendations.Bybalancingthe

threekeyfactors—similarity,genrealignment,andratings—thesystem

ensuresusersreceivetailoredsuggestionsthatcloselymatchtheir

preferencesandgameplayhistory.

Asignificantstrengthofthesystemliesinitsexplanationmechanism.Each

recommendationincludesadetailedbreakdownofthesimilarityscores,genre

overlaps,andotherfactorsinfluencingthecombinedscore.Thistransparency

notonlyenhancesusertrustbutalsoempowersuserstounderstandhow theirchoicesimpacttheresults.Forinstance,ausercanseehowagamelike "Inside"contributedtotherecommendationof"Assassin’sCreedIV:Black Flag"duetosharedgameplaythemesandgenres.

Additionally,theuseofadvancedtechniquessuchasTF-IDFfortext vectorizationandSVDfordimensionalityreductionhasoptimizedthesystem's performance,particularlyinhandlinglargedatasets.Theabilitytoefficiently processandrecommendgamesfromadatasetexceeding15,000records highlightsthesystem’sscalability.Theseinsightsaffirmthesystem’s robustnessanditsabilitytodeliverhigh-qualityrecommendationsinrealworldscenarios.

8.4.2.PlannedEnhancements:

Whilethesystemperformswellinitscurrentstate,severalenhancementsare plannedtoimproveitsfunctionality,accuracy,anduserexperience:

1.DynamicWeightingSystem:Oneofthekeyimprovementswillbe introducingadynamicweightingsystem.Thisfeaturewillallowusersto prioritizefactorssuchasratings,similarity,orgenrealignmentwhen generatingrecommendations.Forexample,auserwhovalueshigh-rated gamesovergenrealignmentcouldadjusttheweightingtofavorratings, leadingtorecommendationsthatalignwiththeirspecificpriorities.This customizationwillfurtherpersonalizetheexperienceandenhanceuser satisfaction. 2.Real-TimeFeedbackCollection:Incorporatingafeedbackmechanismwill allowuserstoratetherelevanceandusefulnessoftherecommendedgames. Thisfeedbackwillbecollectedinrealtimeandintegratedintothe recommendationalgorithm.Byusingcollaborativefilteringtechniques,the

systemcanimproveitsaccuracyandadapttouserpreferencesovertime.For instance,ifusersconsistentlypreferrecommendationswithhighergenre overlap,thesystemcanlearntoweighgenrealignmentmoreheavily.

3.ExpandedVisualizationTools:Visualizationtoolswillbeaddedtohelp usersexploretherecommendationprocessandgaindeeperinsightsintotheir preferences.Forinstance,genredistributionheatmapscouldvisualizehowa user’sselectedgenrescomparetothoseofrecommendedgames.Similarly, similaritygraphscouldillustratetherelationshipsbetweenuser-selected gamesandrecommendedtitles.Thesetoolswillmaketherecommendation processmoreinteractiveandengaging.

4.DiverseDataIntegration:Currently,thesystemreliesontextual descriptions,ratings,andgenres.Futureenhancementswillintegratemore diversedata,suchasuserreviews,gameplaytime,andin-game achievements.Theseadditionalfactorscanhelprefinetherecommendations andmakethemevenmorepersonalized. 5.SupportforMulti-UserRecommendations:Thesystemcanbe enhancedtosupportrecommendationsforgroupsofusers,suchasfriendsor familymemberswhoplaygamestogether.Byanalyzingthepreferencesand gamehistoriesofmultipleusers,thesystemcouldgeneratesuggestionsthat appealtoeveryoneinthegroup.Thisfeaturewillbeparticularlyvaluablefor multiplayergamesandsharedgamingexperiences. 6.Cross-PlatformRecommendations:Toincreaseitsapplicability,the systemcouldrecommendgamesbasedonplatformpreferences,suchasPC, console,ormobile.Userswhopredominantlyplayonaspecificplatform wouldreceiverecommendationstailoredtothatenvironment. 9.Discussion:

9.1.ComparisonandEvaluation:

TheGameRecommenderSystemperformedasexpectedindelivering personalizedrecommendationsbasedonuserinputssuchasage,played games,andpreferredgenres.Theanalysisoftheresultsshowsthatthe systemeffectivelyidentifiesgamesthatalignwithuserpreferences.For instance,therecommendationsprovidedintestingcloselymatchedthe genrealignmentandgameplaypatternsofselectedgames,withsimilarity

scoresandcombinedscoresreflectinglogicalandconsistentrelationships.
Thesystemexceededexpectationsintermsofscalabilityand explainability.Despiteprocessingadatasetofover15,000entries,it consistentlydeliveredrecommendationswithminimallatency.This efficiencycanbeattributedtotheoptimizationtechniquesappliedduring thepreprocessingphase,suchasdimensionalityreductionviaSVDand clusteringtounifysimilargameversions.
However,thesystem'sperformanceshowedslightdeviationsinedge cases.Forexample,whenusersselectednichegameswithlowdata representation,therecommendationssometimesskewedtowardsmore populargameswithinthesamegenre.Thisbehavior,whilelogical, suggestsaneedforimprovementinhandlingsparsedatascenariosto ensurediversityinrecommendations.

Page46of50 9.2.CriticalAnalysis: 9.2.1.ChallengesFacedDuringImplementation:
1.DataPreprocessing:Cleaningandunifyingthedatasetwasoneofthe mostsignificantchallenges.Therawdatasetcontainedinconsistencies suchasduplicategameversions,incompletedescriptions,and overlappinggenrelabels.Implementingaclusteringapproachusing DBSCANtounifygamenamesrequiredextensivefine-tuningtobalance betweenmergingsimilarentriesandpreservinguniquedatapoints.
2.DimensionalityReduction:IntegratingTF-IDFandSVDtoreduce dimensionalityandcomputeitemsimilarityrequiredcarefulparameter tuning.Over-reductionindimensionsriskedlosingimportantsemantic details,whileunder-reductionledtoinefficienciesincomputation.
3.HandlingSparseInputs:Userswithsparseorhighlyspecific preferencesposedachallengetotherecommendationprocess.For example,whenauserselectedfivehighlynichegames,thesystem

sometimesstruggledtoprovidediverserecommendationsduetolimited

overlapingenresordescriptions. 4.Age-BasedFiltering:Designinganage-

basedfilteringmechanism involvedbalancinginclusivitywithappropriateness.Ensuringthat

recommendationsforusersbelow15adheredtotheagerestriction

withoutoverlynarrowingtheresultsrequiredadditionallogicduring recommendationgeneration.

5.Explainability:Creatingclear,interpretableexplanationsfor

recommendationswasbothcriticalandchallenging.Thesystemhadto

presentabreakdownoffactorslikesimilarityandgenrealignmentwithout

overwhelminguserswithexcessivedetail. 9.2.2.SystemPerformanceEvaluation:

Overall,thesystemmetitsobjectivesofdeliveringaccurateandexplainable

recommendations.However,theevaluationhighlightedsomeareasfor improvement:

PrecisioninSparseData:Recommendationsfornichepreferences lackedvariety.

DynamicResponsiveness:Theweightingofsimilarity,genreoverlap,

andratingswasstatic,limitinguser-specificflexibility. 9.3.Enhancements:

Thechallengesidentifiedduringimplementationformthebasisfor

actionablefutureenhancements.Theseinclude:

1.ImprovedHandlingofSparseData:Incorporateahybrid

recommendationapproachthatcombinescontent-basedfilteringwith

collaborativefiltering.Thiswouldleverageuserinteractiondatato

enhancerecommendationswhencontent-basedsimilaritiesare insufficient.

2.DynamicWeighting:Allowuserstocustomizetheweightingoffactors

likesimilarity,genrealignment,andratings.Implementingasimplesliderbasedinterfaceintheappl

icationwouldempoweruserstoprioritizetheir preferencesdynamically.

3.EnhancedDiversityinRecommendations:Introducediversityenhancingalgorithmsthatensur

eabroaderrangeofrecommendations,

evenforuserswithnichepreferences.Forexample,thesystemcouldcap thenumberofrecommendationsfromthesamegenreorintroduce randomnesstolow-rankingrecommendations. 4.RefinedAge-BasedFiltering:Incorporateamorenuancedage-filtering systemthatusesacombinationofageratingsandcontenttagstoprovide appropriateyetdiverserecommendationsforyoungerusers.

5.FeedbackMechanism:Introduceafeatureforuserstorate recommendations.Theseratingscouldbeincorporatedintocollaborative filteringmodels,allowingthesystemtolearnandimproveiteratively.

6.IntegrationofUserBehaviorData:Expandthedatasettoincludedata pointssuchastimespentongames,completionrates,andreview sentiments.Thesebehavioralmetricswouldenhancethe recommendationprocessbycapturingdeeperinsightsintouser preferences.

7.InteractiveVisualizations:Addvisualelements,suchassimilarity graphsandheatmapsofgenreoverlaps,tohelpusersbetterunderstand therecommendationprocess.Thesetoolswouldnotonlyincrease transparencybutalsoimproveuserengagement.

8.MultilingualSupport:Whilethesystemalreadyidentifiessupported languagesforrecommendedgames,expandingtheinterfacetoprovide recommendationsinauser'spreferredlanguagewouldenhance accessibility. 10.Conclusion: TheGameRecommenderSystemsuccessfullyachieveditsgoalofproviding personalizedandexplainablegamerecommendationsbasedonuserinputs. ThroughtheintegrationofadvancedtechniquessuchasTF-IDFvectorization, dimensionalityreductionviaSVD,andclusteringfordataunification,the systemdeliveredhighlyrelevantsuggestionstailoredtouserpreferences.The inclusionoffactorssuchasagefiltering,similarityscoring,genrealignment, andrating-basedweightingensuredtherecommendationswereappropriate, logical,anddiverse. Oneofthesystem'smostsignificantcontributionsisitsuser-centricdesign, emphasizingtransparencyandtrustthroughdetailedexplanationsof recommendations.Thisfeaturenotonlyenhancesuserconfidencebutalso differentiatesthesystemfromgenericblack-boxapproachesin recommendationengines.Additionally,thesystemdemonstratedscalability byprocessingalargedatasetefficiently,provingitssuitabilityfordeployment inreal-worldscenarios. Despitethechallengesencountered,theprojecthighlightedthepotentialfor leveragingcontent-basedtechniquestocreatemeaningfulandpersonalized experiences.Byaddressinglimitationssuchassparsedatahandlingand incorporatingdynamicweighting,thesystemcanevolvetomeetevenmore diverseuserneeds.Thisprojectunderscoresthegrowingimportanceof recommendationsystemsinthegamingindustry,wherepersonalizationplays acriticalroleinusersatisfactionandretention.

Inconclusion,theprojectnotonlydemonstratedthefeasibilityofbuildingan advancedrecommendationenginebutalsoshowcaseditsrelevancein

moderngamingapplications.Theinsightsgainedandenhancements proposedprovideastrongfoundationforfutureimprovementsand applicationsofthesystem.

11.References: 1.PandasDocumentation:Pythondatamanipulationlibrary. https://pandas.pydata.org/docs/ 2.NumPyDocumentation:NumericalcomputingwithPython. https://numpy.org/doc/ 3.Scikit-learn:Machinelearninglibraryforclustering,TF-IDF,andSVD. https://scikit-learn.org/stable/

4.TensorFlow:Machinelearningframeworkforneuralcollaborativefiltering. https://www.tensorflow.org/ 5.Ester,M.,etal.(1996).Adensity-basedalgorithmfordiscoveringclusters inlargespatialdatabaseswithnoise.ProceedingsofKDD-96. https://www.aaai.org/

6.Manning,C.D.,Raghavan,P.,&Schütze,H.(2008).Introductionto InformationRetrieval.CambridgeUniversityPress. DOI:10.1017/CBO9780511809071

7.Strang,G.(2009).IntroductiontoLinearAlgebra.Wellesley-Cambridge Press.

8.He,X.,etal.(2017).NeuralCollaborativeFiltering.ProceedingsofWWW '17.DOI:10.1145/3038912.3052569 9.Lops,P.,etal.(2011).ContentbasedRecommenderSystems:Stateof

theArtandTrends.InRecommenderSystemsHandbook,Springer.

DOI:10.1007/978-0-387-85820-3_3

10.Burke,R.(2002).HybridRecommenderSystems:Surveyand

Experiments.UserModelingandUser-

AdaptedInteraction,12(4),331370.DOI:10.1023/A:1021240730564

11.Su,X.,&Khoshgoftaar,T.M.(2009).ASurveyofCollaborativeFiltering

Techniques.AdvancesinArtificialIntelligence,ArticleID421425. DOI:10.1155/2009/421425

12.Zhang,S.,etal.(2019).DeepLearningBasedRecommenderSystem:A

SurveyandNewPerspectives.ACMComputingSurveys(CSUR),52(1),

1-38.DOI:10.1145/3285029 13.Kingma,D.P.,&Welling,M.(2014).Auto-

EncodingVariationalBayes. arXiv:1312.6114

14.·Anderson,C.A.,&Dill,K.E.(2000).VideoGamesandAggressive

Thoughts,Feelings,andBehaviorintheLaboratoryandinLife.Journalof

PersonalityandSocialPsychology,78(4),772-790.

15.Yee,N.(2006).TheDemographics,Motivations,andDerived

ExperiencesofUsersofMassivelyMulti-userOnlineGraphical

Environments.PRESENCE:TeleoperatorsandVirtualEnvironments,

15(3),309-329.DOI:10.1162/pres.15.3.309

16.Chen,L.,&Pu,P.(2010).TrustBuildinginRecommenderAgents.

ProceedingsofAAMAS2010.https://aamas.csc.liv.ac.uk/

17.Herlocker,J.L.,etal.(2004).EvaluatingCollaborativeFiltering

RecommenderSystems.ACMTransactionsonInformationSystems

(TOIS),22(1),5-53.DOI:10.1145/963770.963772

18.Shaker,N.,Togelius,J.,&Nelson,M.J.(2016).ProceduralContent

GenerationinGames.Springer.DOI:10.1007/978-3-319-42716-4

19.Sifa,R.,etal.(2015).UserBehaviorinGameAnalytics:AReview.

EntertainmentComputing,13,10-22.DOI:10.1016/j.entcom.2015.06.002

20.Bostock,M.,Ogievetsky,V.,&Heer,J.(2011).D3Data-Driven

Documents.IEEETransactionsonVisualizationandComputerGraphics,

17(12),2301-2309.DOI:10.1109/TVCG.2011.185 ·

Sources

1	https://quizlet.com/893436597/elementary-statistical-reasoning-module-2-flash-cards/ INTERNET 1%

EXCLUDE CUSTOM MATCHES	ON
EXCLUDE QUOTES	OFF
EXCLUDE BIBLIOGRAPHY	OFF