

final

by Jonathan Lesmond

Submission date: 18-Oct-2025 02:45PM (UTC+0100)

Submission ID: 265660919

File name: 49131_Jonathan_Lesmond_final_695044_2042574525.docx (148.8K)

Word count: 1644

Character count: 11967



Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

**'Comparison of Long Term Projection of the Welsh population in
Multiple Machine Learning Models'**

Independent Project (ACCA6020)

Jonathan Lesmond (2300728)

BSc (Hons) Applied Computing

**Preferred Supervisor(s): Dr Seena Joseph, Gaynor Thomas, Divya
Chidambaram**

Table of Contents

1. Introduction 3

2. Background to the Proposed Research..... 4

 2.1 Predictive Analytics (PA) and Machine Learning (ML) 4

 2.2 Geographic Information Systems (GIS)..... 4

3. Project Design 5

 3.1. Research Aim & Objectives 5

 3.2 Methods 5

 Data Collection and Sources..... 5

 Technologies 5

 3.3 Project Timeline..... 6

Table of Figures

Figure 1 - Proposed Timeline for the Project..... 6

1. Introduction

According to the StatsWales which is a service owned by the Welsh government to publish datasets to the public, the most recent data that showed population projections by local authority were based on 2018's population estimate.[1] although there are more recent projections of the UK population that include Wales such as the 2022 based projection from the Office for National Statistics (ONS), it does not provide details on specific Welsh local authority populations, section 8 of that publication also states that the ONS are planning to release a 2022 based projection on local authorities in England only.[2]

Now that there is data from the Welsh government providing population estimates from 2019-2024,[3][4] this project aims to give an updated version of population projection which would help give a more accurate depiction of the Welsh Population in the future, it will also evaluate the accuracy of the 2018 based projections. This report will outline the motivation behind the project, the objectives, the methods and tools that will be utilised to achieve those objectives and the justification of the choices.

2. Background to the Proposed Research

2.1 Predictive Analytics (PA) and Machine Learning (ML)

An early account of PA in the 1980s mentions an American company called the Fair Isaac Company (FICO) using mathematical equations to calculate the likelihood that a person will pay back their loan, this likelihood was represented as a numerical score which was named the FICO score, the increasing use of the FICO score in other companies may have inspired how companies outside of the USA calculate credit scores.[5][6]

PA branches from data analytics and focuses on leveraging historical data to make predictions for the future, in more recent years these predictions have increasingly incorporated ML.[7] PA has a lot of applications in fields such as healthcare where monitoring devices worn by patients incorporate algorithms to help staff be aware of risks before an incident occurs.[8] Transportation is another application of PA, for example, the London North Eastern Railway (LNER) who are using ML models trained from data on train journey performance, quantity of passengers, weather conditions etc. to predict delays in train journeys.[9]

Governments use PA and ML in a variety of departments, for example, the NHS uses ML to predict patient demands across hospitals to inform resource allocation.[10] The ONS in one of their publications about national population projections mentioned that they produce these reports to be given to the governments of the constituent countries of the UK to assist with planning in education, health, pensions and economic forecasts.[2]

2.2 Geographic Information Systems (GIS)

GISs allows the visualisation, management and analysis of spatial data,[11][12] GISs are usually used to study a population and their characteristics in specific areas, also known as geodemography.[13] GISs are also used to understand the geographical characteristics or infrastructure of an area.

The origins of GISs date to the 1960s when the first GIS, the Canada Geographic Information System (CGIS)[11][14] was created in collaboration with IBM for a Canadian government funded program called the Canada Land Inventory to reduce their difficulty managing natural resources due to the size of the country.[15] Esri, a reputed GIS software company was founded using early advancements in technology.[14] from the 1990s to modern day more major advancements in satellite imagery and observation, and hardware made GISs more powerful and popular in everyday life through systems such as Google Maps and Uber. GISs have more frequent usage in environmental sciences and urban design to assist construction projects, research on air quality, predictive analytics on crime etc.[14]

3. Project Design

3.1. Research Aim & Objectives

The aim of the project is to create an updated projection of the Welsh population, including the population of each local authority that will be visualised through a geographic information system, the project also aims to compare the accuracy of predictions from various types of machine learning models.

The key objectives of the project are as follows:

1. To produce an updated projection of the Welsh population by local authority with multiple different machine learning models
2. To produce a visualisation that allows easy understanding of the projections
3. To compare the accuracy of the model's predictions with each other and with historical projection data

3.2 Methods

An experiment will be the primary research method of this project. The reasoning behind this being that project will make evaluations and draw comparisons between different machine learning algorithms and historical data from government publications

Data Collection and Sources

Collecting quantitative data from publicly available repositories is the main method of data collection for this project, the sources that data will be collected will include StatsWales, DataMapsWales, Digital Health and Care Wales etc. The data collected will relate to housing, education, health and social care, mortality, migration etc.

Technologies

The development of the machine learning models will be done in Python, The Python programming language has the most support for machine learning operations through libraries such as Scikit-learn, TensorFlow, Keras etc. Python also has libraries for reading data from CSV or SQL files through polars and pandas, SQLAlchemy is another popular library within Python used for inserting data through Python code.[16]

For storing historical data and the new data that will be produced, MySQL will be used, this is due to its compatibility with Python through MySQL Connector, through polars and MySQL Connector, the MySQL database will allow querying of data via Python code.[17]

Quantum GIS (QGIS) will be the GIS of choice for the project, QGIS is a widely used free and open source GIS used by various authorities around the world for case studies in countries such as Saint Lucia, Nigeria, Portugal etc. [18]

3.3 Project Timeline

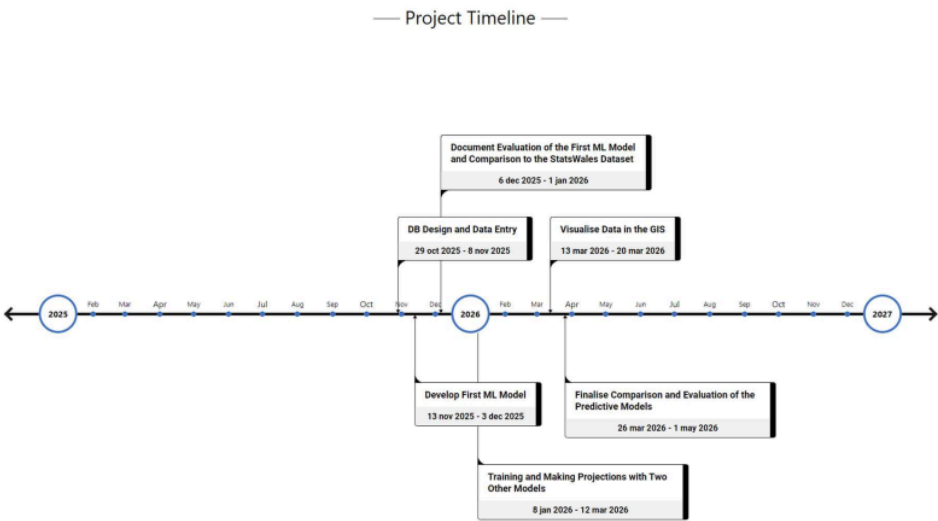


Figure 1 - Proposed Timeline for the Project

References

- [1] Welsh Government. (August). *Population projections by local authority and year*. Available: <https://statswales.gov.wales/Catalogue/Population-and-Migration/Population/Projections/Local-Authority/2018-based/populationprojections-by-localauthority-year>.
- [2] Office for National Statistics. (January). *National population projections: 2022-based*. Available: <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationprojections/bulletins/nationalpopulationprojections/2022based>.
- [3] Office for National Statistics. (Sept 25). *Population estimates by local authority, year, sex and age*. Available: https://stats.gov.wales/en-GB/82d9faea-b515-41cd-aedb-8f4594f66ce5#about_dataset.
- [4] Welsh Government. (July). *Population estimates by local authority and year*. Available: <https://statswales.gov.wales/Catalogue/Population-and-Migration/Population/Estimates/Local-Authority/populationestimates-by-localauthority-year>.
- [5] G. Hagler. G. Hagler. Credit score. Credit score. *EBSCO EBSCO* 2022. Available: <https://www.ebsco.com/research-starters/business-and-management/credit-score>.
- [6] R. Sheposh. R. Sheposh. Predictive analytics. Predictive analytics. *EBSCO EBSCO* 2024. Available: <https://www.ebsco.com/research-starters/science/predictive-analytics#full-article>.
- [7] P. Ongsulee, V. Chotchaung, E. Bamrungsi and T. Rodcheewit. P. Ongsulee, V. Chotchaung, E. Bamrungsi and T. Rodcheewit. Big data, predictive analytics and machine learning. Big data, predictive analytics and machine learning. Presented at 2018 16th International Conference on ICT and Knowledge Engineering (ICT&KE). Presented at 2018 16th International Conference on ICT and Knowledge Engineering (ICT&KE). 2018, . DOI: 10.1109/ICTKE.2018.8612393.
- [8] M. S. Supriya and K. Sasidaran. M. S. Supriya and K. Sasidaran. "Machine learning for big data," in *Applications of Big Data in Large- and Small-Scale Systems*, S. Goundar and P. K. Rani, Eds. 2021, Available: <https://services-igi-global-com.ezproxy.uwtsd.ac.uk/resolvedoi/resolve.aspx?doi=10.4018/978-1-7998-6673-2.ch004>. DOI: 10.4018/978-1-7998-6673-2.ch004.
- [9] London North Eastern Railway (LNER). London North Eastern Railway (LNER). The tech keeping A watch on train times. The tech keeping A watch on train times. 2025. Available: <https://www.lner.co.uk/news/the-tech-keeping-a-watch-on-train-times/>.
- [10] I. P. S. Sethi, O. P. Gupta, R. K. Dwivedi, M. Kapoor, S. Bhaisare, V. K. Chaudhari and S. Varma. I. P. S. Sethi, O. P. Gupta, R. K. Dwivedi, M. Kapoor, S. Bhaisare, V. K. Chaudhari and S. Varma. Machine learning in government applications: A review. Machine learning in government applications: A review. *Procedia Computer Science Procedia Computer Science* 258258pp. 1365–1371. pp. 1365–1371. 2025. Available: <https://www.sciencedirect-com.ezproxy.uwtsd.ac.uk/science/article/pii/S1877050925014711>. DOI: 10.1016/j.procs.2025.04.369.
- [11] X. Li, J. Yue, S. Wang, Y. Luo, S. Cheng, J. Zhou, D. Xu and H. Lu. X. Li, J. Yue, S. Wang, Y. Luo, S. Cheng, J. Zhou, D. Xu and H. Lu. Development of geographic information system

architecture feature analysis and evolution trend research. Development of geographic information system architecture feature analysis and evolution trend research. *Sustainability* 1616(1), 2024. Available: <https://www.proquest.com/scholarly-journals/development-geographic-information-system/docview/2912828553/se-2?accountid=130472> https://libkey.io/libraries/1422/openurl?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rft_id=info:sid/ProQuest&rft_val_fmt=journal&rft.genre=article&rft.jtitle=Sustainability&rft.atitle=Development+of+Geographic+Information+System+Architecture+Feature+Analysis+and+Evolution+Trend+Research&rft.au=Li%252C+Xiao%253BYue%252C+Jianwei%253BWang%252C+Shao-hua%253BLuo%252C+Yifei%253BCheng%252C+Su%253BZhou%252C+Junyuan%253BXu%252C+Dachuan%253BLu%252C+Hao.&rft.aulast=Li&rft.aufirst=Xiao&rft.date=2024-01-01&rft.volume=16&rft.issue=1&rft.spage=137&rft.isbn=&rft.btitle=&rft.title=Sustainability&rft.issn=&rft_id=info:doi/10.3390%252Fsu16010137 <https://doi.org/10.3390/su16010137>. DOI: 10.3390/su16010137.

[12] A. L. Franklin, T. Mott and T. H. L. Williams. A. L. Franklin, T. Mott and T. H. L. Williams. Coproduction in the U.S. department of defense: Examining how the evolution of geographic information systems (GIS) expands non-traditional partner engagement. Coproduction in the U.S. department of defense: Examining how the evolution of geographic information systems (GIS) expands non-traditional partner engagement. *Policy & Internet* 55(4), pp. 387–401. pp. 387–401. 2013. Available: <https://doi.org/10.1002/1944-2866.POI345>. DOI: 10.1002/1944-2866.POI345.

[13] A. Hetmanenko, V. Khomutov, O. Nikolaieva and S. Kudelina. A. Hetmanenko, V. Khomutov, O. Nikolaieva and S. Kudelina. Geodemography as a component of population geography. Geodemography as a component of population geography. *J Educ Health Sport J Educ Health Sport* 1313(5), pp. 161. pp. 161. 2023. . DOI: 10.12775/jehs.2023.13.05.021.

[14] R. Steenson. R. Steenson. The history of geographic information systems (GIS). The history of geographic information systems (GIS). 2019. Available: <https://www.bcs.org/articles-opinion-and-research/the-history-of-geographic-information-systems-gis/>.

[15] M. M. Castron. M. M. Castron. Colonialism, computerized the canada land inventory and the canada geographic information system at library and archives canada. Colonialism, computerized the canada land inventory and the canada geographic information system at library and archives canada. *Archivaria* 93932022. Available: <https://www.proquest.com/scholarly-journals/colonialism-computerized-canada-land-inventory/docview/2770087481/se-2?accountid=130472> https://libkey.io/libraries/1422/openurl?ctx_ver=Z39.88-2004&ctx_enc=info:ofi/enc:UTF-8&rft_id=info:sid/ProQuest&rft_val_fmt=journal&rft.genre=article&rft.jtitle=Archivaria&rft.atitle=Colonialism%252C+Computerized+The+Canada+Land+Inventory+and+the+Canada+Geographic+Information+System+at+Library+and+Archives+Canada&rft.au=Castron%252C+Melissa+M.&rft.aulast=Castron&rft.aufirst=Melissa&rft.date=2022-04-01&rft.volume=93&rft.issue=&rft.spage=136&rft.isbn=&rft.btitle=&rft.title=Archivaria&rft.issn=03186954&rft_id=info:doi/.

[16] A. Géron. A. Géron. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow* 2022 Available: <http://ebookcentral.proquest.com/lib/trinity/detail.action?docID=30168989>.

[17] Oracle. *Chapter 1 Introduction to MySQL Connector/Python*. Available: <https://dev.mysql.com/doc/connector-python/en/connector-python-introduction.html>.

[18] QGIS Foundation. *Case Studies*. Available: <https://qgis.org/project/case-studies/>.

FINAL GRADE

GENERAL COMMENTS

Student **Jonathan Lesmond**
Mark **68**

68/100

Intro / A&O (400 words)		
Project title; Identification of a specific problem to be investigated; Well-defined aims and SMART objectives; Explaining why this project is worth doing; Discussion of what the project needs to do in order to be a success; Description any deliverables that will be produced at the end of the project <i>The introduction provided is technical and reads like a literature review than an introduction.</i> <i>Lacks a strong statement of why this project matters for stakeholders (e.g., planners, policymakers).</i> <i>Begin with a broad context: Why population projections are important for Wales (planning, resource allocation, policy).</i> <i>Then narrow down to the gap: Current projections are outdated or incomplete.</i> <i>End with a clear purpose statement: What this project will deliver and why it's valuable.</i>	20	15

<i>Aim is clear and most of the objectives are aligned with the proposed aim, but they are broad so need to be refined.</i>		
Background to the Project / Literature Review (400 words)		
Description of how the work to be done by in the project relates to the work identified in the literature review; Discussion of any ethical issues raised by the project Identifying what has already been done within the field; Topic familiarity; Brief literature review using appropriate sources – 6-8 journal papers and book chapters rather than just websites; Selection of relevant models & concepts <i>Background section shows the basic understanding of the proposed area.</i> <i>To make this section stronger and more relevant to your project:</i> <i>Focus on why these technologies matter for your project rather than giving long historical details.</i> <i>Example: Explain how predictive analytics and GIS will help improve Welsh population projections.</i> <i>Connect the literature to your project choices.</i> <i>Example: If you plan to use certain machine learning</i>	20	14

<i>models, mention that research shows these models work well for population forecasting.</i> <i>Highlight the gap you are addressing.</i> <i>Example: Current Welsh projections are based on older data and lack local detail, your project solves this.</i> <i>Keep it short and clear, focusing on relevance, gaps, and justification</i>		
Project Design (500 words)		
Clear statement of which development methods are to be used and why they are appropriate; Clear statement of which tools are to be used and why they are appropriate; Clear description of how data collected will be analysed <i>project design shows good planning and identifies relevant tools like Python, MySQL, and QGIS</i> <i>Need to provide the details of ML models Instead of saying "multiple ML models," specify which ones you plan to use and why.</i> <i>Define evaluation metrics</i> <i>Explain how you will measure model accuracy.</i> <i>Show the steps from data collection → preprocessing →</i>	30	19

<i>model training → evaluation → GIS visualisation. This makes your design easier to follow.</i> <i>Make the time line realistic:</i> <i>Ensure milestones reflect achievable goals:</i> • <i>Data collection and cleaning</i> • <i>Model development and testing</i> • <i>GIS visualisation</i> • <i>Final evaluation</i>		
Resources / Planning (200 words)		
List of any deliverables; Clear diagram showing both timing and resource use; Completed ethics form; Limited and relevant appendices <i>Good identification of data sources and tools</i> <i>Ethics and GDPR compliance are not mentioned, this is essential when handling government datasets.</i> <i>Timeline is referenced but not clearly visible;</i> <i>include a detailed Gantt chart with milestones for data collection, model development, and evaluation.</i>	15	10

<i>Resource allocation (time for ML model training vs GIS integration) should be realistic.</i>		
Document Presentation		
Correct spelling, grammar and punctuation; Flow of narrative/fluency; Coherence; Word processed ; Correct use of referencing with bibliography; Strict word limit <i>Document is structured well.</i> <i>Citations and references need to be improved.</i> <i>Evidence of spelling and grammatical errors.</i>	15	10
TOTAL	100	68

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9
