**MACHINE LEARNING-POWERED CLIENT RISK PROFILING: AN EXPLAINABLE AI APPROACH FOR RADIANT FINANCIAL**

Jamie Smiles, Newcastle University, UK

The abstract goes here…

1. Introduction

The use of Artificial Intelligence (AI) and Machine Learning (ML) in financial services is steadily reshaping the way client risk is assessed, managed, and communicated. Traditionally, risk profiling within wealth management firms has relied on static questionnaires, subjective adviser input, and simplified scoring mechanisms [1]. This approach, while compliant, often overlooks the wealth of structured and unstructured data available within an organisation's systems. As regulatory bodies such as the Financial Conduct Authority (FCA) advocate for greater transparency and accountability in AI-driven decision-making, there is a pressing need to develop explainable, data-driven methods that enhance risk assessment practices [2].

Radiant Financial, a rapidly growing UK-based financial services consolidator, exemplifies this challenge. Following its acquisition of Seven Bridges Investment Management (SBIM), Radiant has inherited a fragmented data landscape, with legacy systems such as Wave CRM and Intelliflo housing vast but underutilised client data. The absence of integrated, predictive analytics limits the firm's ability to provide personalised, evidence-based risk assessments, impacting both operational efficiency and client trust. The upcoming deployment of a centralised data warehouse further presents a timely opportunity to design and prototype AI-driven risk profiling tools that align with Radiant's strategic goals.

This project aims to develop a machine learning-based client risk profiling model for Radiant Financial, leveraging historical client and plan data from Intelliflo CRM. The focus is twofold: firstly, to replace subjective, questionnaire-based risk assessments with an explainable AI model that transparently evaluates risk factors; and secondly, to provide actionable insights for financial advisers through clustering and segmentation of client profiles. The project will place particular emphasis on the use of Explainable AI (XAI) techniques such as SHAP values to ensure compliance with FCA guidelines and maintain stakeholder trust.

The significance of this work lies at the intersection of business and academia. For Radiant, this model represents a pathway to improved risk assessment accuracy, enhanced client engagement, and data-driven decision-making. From an academic perspective, the project contributes to ongoing research in the application of interpretable machine learning methods within regulated industries, addressing challenges of fairness, transparency, and practical deployment.

The remainder of this dissertation is structured as follows. Section 2 reviews relevant literature on AI in financial services, risk profiling methods, and explainability frameworks. Section 3 details the methodology adopted, including data collection, preprocessing, modelling, and evaluation techniques. Section 4 outlines the work undertaken, highlighting key decisions and implementation challenges. Section 5 presents the results and evaluates their business impact. Section 6 concludes with a reflection on the project's achievements, limitations, and recommendations for future work.

1. Background
2. Methodology
   1. Data Sources and Extraction

The primary data source for this investigation is Intelliflo Office (IO) – a comprehensive CRM and back-office platform widely adopted across the UK financial services sector. IO serves as the operational core for Radiant, supporting client onboarding, financial planning, policy and investment tracking, document generation, and regulatory recordkeeping. It is also positioned internally at Radiant as the organisation’s Single Source of Truth (SSOT) for structured client and product data.

Technically, IO is underpinned by a relational database comprising dozens of interlinked tables (e.g., Client, Plan, Valuations, Tasks, Contact etc.). These tables contain highly granular information – for instance, contact details are stored in a single value field that can represent mobile numbers, email addresses, or postal codes, depending on metadata **(Appendix A)**. While flexible, this weak typing results in poor input validation and prevents efficient querying (e.g., the proportion of clients with invalid or missing postcodes is not natively reportable). Such limitations are exacerbated by the large scale of Radiant’s client base, which exceeds 19,991 individuals, and the amount of active policies they hold, which exceeds 60,390 plans.

To mitigate these data quality gaps and enable near-real-time insights, IO connects via API to major provider platforms such as Transact, Nucleus, Aviva, and AJBell. These integrations stream in holdings, cash movements, valuations, and fees. The ambition is to consolidate these disparate sources into a coherent framework that enables comprehensive portfolio oversight across platforms allowing for management information (MI) reporting on Assets Under Administration (AUA). However, platform fragmentation creates challenges in producing a unified and timely view, particularly around transactional data (e.g. contributions, withdrawals, fees, or switches) with client-level risk and financial objectives. Radiant acknowledge these issues and are actively engaged in data warehousing programmes, where provider APIs and IO are being unified to produce vertically integrated dashboards and support FCA-mandated reporting[[1]](#footnote-1).

The second data source employed was extracted from Dynamic Planner (DyP) – a psychometric risk profiling tool used by financial advisers across Radiant to assess clients’ investment risk preferences. DyP generates risk ratings such as Attitude to Risk (ATR) and Capacity for Loss (CFL) by interpreting structured questionnaire responses. These outputs are used to guide client recommendations in line with FCA suitability expectations. These scores support recommendation suitability and are routinely captured during adviser-client reviews.

However, these outputs are not deterministic. An ATR score is not a standalone decision driver, nor does it prescribe a specific portfolio. Rather, it acts as an informative input to a broader advisory conversation. For instance, a client with a DyP rating of 5 may ultimately choose a less aggressive portfolio based on personal comfort or current life circumstances. As such, the purpose of this machine learning investigation is not to replace psychometric assessment or adviser discretion, but to augment it – offering an explainable insight into likely client behaviours. From an operational perspective, such tools allow Radiant to:

* Pre-empt risk trends across unprofiled clients (prior to ATR completion)
* Set internal triggers for expected changes (e.g. entering retirement)
* Support adviser-client conversations with contextualised, data-driven framing

In practice, both IO and DyP do not expose direct database access. Instead, structured exports were retrieved via embedded reporting interfaces and downloaded as .csv files. This restricted the ability to perform joined queries or schema-level analysis. Reports from both systems were fragmented, partially duplicated, and loosely structured – for example, free-text fields often lacked formatting, column names varied between exports, and row duplications were frequent. As a result, significant manual preprocessing was required to de-duplicate entries, normalise formatting, and extract categorical values from inconsistent text fields. This preparatory work preceded the machine learning investigation and formed the foundation of the revised datasets used throughout the project.

* 1. Data Merging and Preprocessing

Following initial exploratory analysis, a comprehensive preprocessing pipeline was developed to clean, deduplicate, and engineer features across the five raw datasets (Table 1). This stage aimed to address issues of inconsistency, missing values, anonymisation, and semantic duplication, while enhancing model readiness and interpretability for downstream risk profiling and segmentation.

Table 1: Raw Input Summary

| **File Name** | **Description** | **Rows** | **Cols** | **Example Fields** |
| --- | --- | --- | --- | --- |
| CLMM - Multiple Plans.csv | Exported extract of client and policy-level data from IntelliFlow Office. Captures demographic, contact, adviser, plan, provider and vulnerability attributes. | 140,138 | 83 | DOB, PolicyStatus, PlanType, ReviewDueDate |
| Client Earnings Report.csv | Financial income report derived from IO, providing gross and net income breakdowns, affordability scores, salary, and earnings metadata. | 2,450 | 35 | Salary, Annual Earnings, Disposable Income, Smoker |
| VulnerableClientsReport.csv | CRM-level vulnerability flagging report, capturing both structured and free-text assessments related to client support needs. | 6,248 | 26 | VulnerabilityType, DateAssessed |
| Client Types.csv | Categorical segmentation of clients by entity type, gender, employment status, nationality, and trust identifiers. | 34,152 | 14 | Client Type, Employment Status, Marital Status |
| Attitude to Risk Report.csv | Dynamic Planner export of psychometric testing outcomes for client risk profiling; includes selected and system-generated scores. | 706,149 | 14 | Risk Profile, Generated Risk Profile Number, Risk Notes |

The original exports from IO and DyP were not immediately suitable for analysis due to high volumes of missing or unstructured data, redundant or duplicated columns, including both semantic overlaps (e.g. ClientRef and ClientMigrationRef) and format variations (e.g. duplicated client IDs), inclusion of personally identifiable information (PII) across multiple columns (e.g. names, NI numbers, email addresses) requiring systematic removal or anonymisation, and structural inconsistencies (e.g. client-level data nested in plan-level rows, or clients appearing multiple times across datasets with conflicting data entries).

| **File Name** | **Description** | **Rows** | **Cols** | **Example Fields** |
| --- | --- | --- | --- | --- |
| plansRevised.xlsx | As above | 60,390 | 21 | AdviserType, DOB, Nationality, ClientRef, ClientMigrationRef |
| earningsRevised.xlsx | As above | 2,450 | 22 | Marital Status, Employment Status, Occupation |
| vulnRevised.xlsx | As above | 6,248 | 14 | Flag\_FinancialStruggle, Flag\_MentalHealth |
| typesRevised.xlsx | As above | 32,175 | 5 | Client Type, Employment Status, Marital Status |
| revisedRiskScores.xlsx | As above | 10,875 | 3 | Client.Id, Has Risk Score, FinalATR |

**ADDITIONS**

**FCA Standards – Examples of Extensions**

If you ever want to deepen regulatory framing, here’s what you can cite:

1. **FCA FG21/1: Guidance on Fair Treatment of Vulnerable Customers**
   * You could mention this when explaining why vulnerability flags were preserved and engineered into the dataset.
2. **MiFID II / PROD 3 / COBS 9 Requirements**:
   * These require firms to demonstrate *suitability* and *appropriateness* when making investment recommendations.
   * Your ML project indirectly supports this by improving pre-review understanding of a client’s risk profile.
3. **Dynamic Planner’s own documentation**:
   * If you have access to DyP docs (e.g. how they score ATR from 1 to 10), that could justify use of those ranges in modelling and interpretation.

You don’t need all of this in 3.1, but it’s useful when writing your **Background** or **Evaluation** chapters — especially if the goal is to show potential business impact and alignment with industry expectations.

1. Results and Evaluation
2. Conclusion
3. Recommendations

Acknowledgements

References

[1] Dynamic Planner, ‘About Dynamic Planner’s Risk Profiling System’, Dynamic Planner. Accessed: May 12, 2025. [Online]. Available: https://dynamicplanner.com/risk-sustainability-profiling/

[2] Harmonic Finance & Operations, ‘Disrupting Tradition: How AI Is Reshaping Wealth Management In The UK’. Accessed: May 12, 2025. [Online]. Available: https://harmonicfinance.com/insight/disrupting-tradition-how-ai-is-reshaping-wealth-management-in-the-uk

[3] S. Barocas, M. Hardt, and A. Narayanan, ‘Fairness in Machine Learning’, 2020.

A  APPENDICES

In the appendix section, three levels of Appendix headings are available.

Fix the Appendix List style to A.X….

A.1 Part One

1. ANOTHER SECTION

The next subsections provide instructions on how to insert figures, tables, and equations in your document.

* 1. Tables

Tables are “float elements” which should be inserted after their first text reference and have specific styles for identification. Do not use images to present tables, or they will be inaccessible to readers using assistive technologies.

Authors can insert tables by using the MS Word option (INSERT ->Table) and providing the required row and column size. Every table must have a caption (title) above it, which must have the **“TableCaption**” style applied. Please note that tables **should not** be supplied as[[2]](#footnote-2) image files, but if they are images they must have the “Image” style applied. As an example, Table 1 shows all the styles available in this template, to be applied to the respective element of your text.

Table 1: Styles available in the Word template

| Style Tag | Definition | Style Tag | Definition |
| --- | --- | --- | --- |
| Title\_document | main title of article | ListParagraph | list items |
| Subtitle | subtitle of article | Statements | math statements |
| Authors | author name | Extract | block quotations |
| Affiliation | author affiliation information | Algorithm Caption | caption for algorithm |
| AuthNotes | footnote to author(s) | AckHead | heading for acknowledgements |
| Abstract | abstract text | AckPara | acknowledgements text |
| CCSHead | heading for CSS Concepts | GrantSponsor | sponsor of grant |
| CCSDescription | CSS terms | GrantNumber | number for the grant |
| KeyWordHead | heading for keywords | ReferenceHead | heading for references |
| Keywords | keywords text | Bib\_entry | references |
| ORCID | author's ORCID # | AppendixH1 | appendix heading level 1 |

Tables can be very difficult for people using screen reader technology to understand unless they include markup that explicitly defines the relationships between all the parts (i.e.: headers and data cells). *A key to making data tables accessible to screen reader users is to clearly identify column and row headers.* In Word, authors should identify which row or rows contain column headers. Below are the steps to do this:

1. Select that table’s row, then right-click the row and select “Table Properties”;
2. In the *Table Properties* window, click the *Row* tab and select the box that says “Repeat as header row at the top of each page.”

Or

Apply the “table head” style by highlighting the respective row and applying the “**TableHead**” style found in the “Body Element” section of the ACM Master Article Template.

* 1. Figures

Figures are “float elements” which should be inserted after their first text reference, and have specific styles for identification. Insert a figure and apply the “**Image**” paragraph style to it. For the figure caption, apply the style “**FigureCaption.**”

To accommodate readers with color vision differences, figures should still be usable when printed in grayscale. Refer to elements of the figure with non-color terms, for example “indicated as squares” instead of “indicated in blue”. Use different patterns in bar charts, different line patterns in graphs, and different shapes in plots to distinguish groups of elements and reinforce color differences.



Figure 1: 1907 Franklin Model D roadster. Photograph by Harris & Ewing, Inc. [Public domain], via Wikimedia Commons. (https://goo.gl/VLCRBB)

* 1. Equations

There are two types of math equations: the *numbered display math equation* and the *un-numbered display math equation*. Below are examples of both.

* + 1. *DisplayFormula.*

The **DisplayFormula** style is applied in the numbered math equation. A numbered display equation always has an equation number (label) on the right.

(1)

* + 1. DisplayFormula.Unnum*.*

The **DisplayFormulaUnnum** style is applied only in unnumbered equations. An unnumbered display equation never contains an equation number Bertot and Grimes (2012) on the right—this element distinguishes it from the numbered equation.

Please note: the subsequent text after the **DisplayFormula** (numbered equation) or **DisplayFormulaUnnum** (unnumbered equation) must have the paragraph style **ParaContinue** applied.

* 1. Math statements

Math statements should have the “Statement” style applied.

**Theorem/Proof/Lemma.** Math statements should have the “**Statement**” style applied. This paragraph is an example of the “**Statement**” style.

1. Citing Related Work

This section cites a variety of journal [5, 15], conference [1, 6, 8, 12, 13], and magazine [3] articles to illustrate how they appear in the references section. It also cites books [9, 10], a technical report [7], a PhD dissertation [4], an online reference [14], a software artifact [11], and a dataset [2].

ACKNOWLEDGMENTS

Identification of funding sources and other support, and thanks to individuals and groups that assisted in the research and the preparation of the work should be included in an acknowledgment section, which is placed just before the reference section in your document.

REFERENCES

1. Atul Adya, Paramvir Bahl, Jitendra Padhye, Alec Wolman, and Lidong Zhou. 2004. A multi-radio unification protocol for IEEE 802.11 wireless networks. In Proceedings of the IEEE 1st International Conference on Broadnets Networks (BroadNets’04) . IEEE, Los Alamitos, CA, 210–217. https://doi.org/10.1109/BROADNETS.2004.8
2. Sam Anzaroot and Andrew McCallum. 2013. UMass Citation Field Extraction Dataset. Retrieved May 27, 2019 from <http://www.iesl.cs.umass.edu/data/data-umasscitationfield>
3. Martin A. Fischler and Robert C. Bolles. 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM 24, 6 (June 1981), 381–395. https://doi.org/10.1145/358669.358692
4. Chelsea Finn. 2018. Learning to Learn with Gradients. PhD Thesis, EECS Department, University of Berkeley.
5. Jon M. Kleinberg. 1999. Authoritative sources in a hyperlinked environment. J. ACM 46, 5 (September 1999), 604–632. https://doi.org/10.1145/324133.324140
6. Matthew Van Gundy, Davide Balzarotti, and Giovanni Vigna. 2007. Catch me, if you can: Evading network signatures with web-based polymorphic worms. In Proceedings of the first USENIX workshop on Offensive Technologies (WOOT ’07) . USENIX Association, Berkley, CA, Article 7, 9 pages.
7. James W. Demmel, Yozo Hida, William Kahan, Xiaoye S. Li, Soni Mukherjee, and Jason Riedy. 2005. Error Bounds from Extra Precise Iterative Refinement. Technical Report No. UCB/CSD-04-1344. University of California, Berkeley.
8. David Harel. 1979. First-Order Dynamic Logic. Lecture Notes in Computer Science, Vol. 68. Springer-Verlag, New York, NY. <https://doi.org/10.1007/3-540-09237-4>
9. Jason Jerald. 2015. The VR Book: Human-Centered Design for Virtual Reality. Association for Computing Machinery and Morgan & Claypool.
10. Prokop, Emily. 2018. The Story Behind. Mango Publishing Group. Florida, USA.
11. R Core Team. 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/
12. Brian K. Reid. 1980. A high-level approach to computer document formatting. In Proceedings of the 7th Annual Symposium on Principles of Programming Languages. ACM, New York, 24–31. <https://doi.org/10.1145/567446.567449>
13. John R. Smith and Shih-Fu Chang. 1997. Visual Seek: a fully automated content-based image query system. In Proceedings of the fourth ACM international conference on Multimedia (MULTIMEDIA ’96). Association for Computing Machinery, New York, NY, USA, 87–98. https://doi.org/10.1145/244130.244151
14. TUG 2017. Institutional members of the LaTeX Users Group. Retrieved May 27, 2017 from <http://wwtug.org/instmem.html>
15. Alper Yilmaz, Omar Javed, and Mubarak Shah. 2006. Object tracking: A survey. ACM Comput. Surv. 38, 4 (December 2006), 13–es. https://doi.org/10.1145/1177352.1177355

A  APPENDICES

In the appendix section, three levels of Appendix headings are available.

A.1 Part One

1. Save as you go and backup your file regularly.
2. Do not work on files that are saved in a cloud directory. To avoid problems such as MS Word crashing, please only work on files that are saved locally on your machine.
3. Equations should be created with the built-in Microsoft® Equation Editor included with your version of Word. (Please check the compatibility at <http://tinyurl.com/lzny753> for using MathType.)
4. Please save all files in DOCX format, as the DOC format is only supported for the Mac 2011 version.
5. Tables should be created with Word’s “Insert Table” tool and placed within your document. (Tables created with spaces or tabs will have problems being properly typeset. To ensure your table is published correctly, Word’s table tool must be used.)
6. Do not copy-and-paste elements into the submission document from Excel such as charts and tables.
7. Footnotes should be inserted using Word’s “Insert Footnote” feature.
8. Do not use Word’s “Insert Shape” function to create diagrams, etc.
9. Do not have references appear in a table/cells format as it will produce an error during the layout generation process.
10. MS Word does not consistently allow the original formatting to be modified in the text. In these cases, it is best to copy all the document’s text from the specific file and paste into a new MS Word document and then save it.
11. At times there are font problems such as “odd” stuff/junk characters that appear in the text, usually in the references. This can be caused by a variety of reasons such as copying-and-pasting from another file, file transfers, etc. Please review your text prior to submission to make sure it reads correctly.

A.2 Placeholder Text

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Vulputate sapien nec sagittis aliquam. Malesuada fames ac turpis egestas sed tempus urna. Posuere sollicitudin aliquam ultrices sagittis orci. Consequat id porta nibh venenatis cras sed felis eget.

1. While this Capstone project does not directly resolve these systemic AUA reconciliation challenges, it is situated within the same operational landscape. Thus, decisions regarding data selection, preparation, and interpretation are made with an awareness of the broader architectural and regulatory constraints. [↑](#footnote-ref-1)
2. Author’s address: Your name, [your\_email@ncl.ac.uk](mailto:your_email@ncl.ac.uk), Newcastle University, Newcastle upon Tyne, UK [↑](#footnote-ref-2)