Applied Journal of Economics, Law and Governance https://doi.org/10.57017/ajelg.v2.i1(3).07

Analysing the Determinants of Graduate Unemployment in Tunisia Using Machine Learning

Sami MESTIRI ✉
Faculty of Management and Economic Sciences of Mahdia
University of Monastir, Tunisia
https://orcid.org/0000-0002-2060-3242

Abstract

This paper analyses the determinants of youth graduate unemployment in Tunisia by combining classical econometric methods (logistic regression) with three machine learning algorithms (Random Forest, XGBoost, RBF-kernel SVM) applied to an original survey of 1,200 Tunisian graduates. The econometric results reveal that female gender, belonging to the engineering field, and education-employment mismatch both vertical (overqualification) and horizontal (field misalignment) are the most significant determinants. The machine learning analysis confirms the predominance of gender and uncovers non-linear interactions: the protective effect of engineering is significantly attenuated for women, revealing that the gender gap persists even in high-demand fields. A feature importance plot derived from XGBoost quantifies each variable's contribution, making the transition from predictive modelling to policy governance more transparent. XGBoost and SVM offer the best predictive performance, outperforming logistic regression on F1-score and AUC-ROC. These findings call for targeted policies against gender discrimination, differentiated reform of the university curriculum, and improved recruitment transparency.

Keywords: graduate youth unemployment; logistic regression; random forests; XGBoost; education-employment mismatch; human capital.

JEL Classification: J64; J24; C25; C45.

Introduction

For two decades, Tunisia has exhibited a remarkable socio-economic paradox: despite a spectacular expansion of higher education the enrolment rate increased from 12% in 1990 to over 36% in 2020 graduate youth unemployment has remained chronically high, oscillating between 28% and 35% depending on the year. This phenomenon, referred to as the diploma paradox, reflects a structural disconnect between the supply of university training and the actual needs of the labour market.

The revolution of 14 January 2011 brought to light the demands for employment and dignity from an overqualified yet labour-market-excluded youth. According to the National Institute of Statistics (INS, 2023), the national unemployment rate stood at 15.2% in 2023; however, higher-education graduates recorded a rate of 26.7%, compared with only 4.5% for individuals without qualifications. Our survey of 1,200 graduates confirms and amplifies this diagnosis: the observed unemployment rate reaches 30.9%, with considerable gender disparities.

This study aims to identify and rank the determinants of unemployment by mobilising a dual approach: econometric (logit/probit) and machine learning (Random Forest, XGBoost, SVM). The originality of this work lies in the application of these approaches to a sample of 1,200 graduates, providing sufficient statistical power to detect effects that prior small-sample studies could not identify.

Four hypotheses structure the analysis:

H1: Education-employment mismatch is the primary determinant of unemployment.

H2: Female gender significantly increases the probability of unemployment.

H3: Reliance on personal connections is the most discriminating factor according to machine learning.

H4: The gender gap in unemployment risk persists even within high-demand fields such as engineering.

The remainder of the article is organised as follows: Section 1 presents the theoretical framework. Section 2 describes the data and methodology. Section 3 presents the empirical results, including gender-field interaction analysis and a feature importance plot. Section 4 discusses the policy implications and last section concludes.

1. Theoretical Framework

Human Capital and Signalling Theories

Human capital theory, formalised notably by Becker (1964) and Mincer (1974), rests on the idea that education constitutes a productive investment. It improves skills, raises the marginal productivity of workers, and translates, on the labour market, into higher wages and a lower probability of unemployment. Within this framework, the expected relationship is monotonically decreasing: the higher the level of education, the lower the risk of unemployment.

However, our empirical results suggest a weakening, if not a breakdown, of this mechanism in Tunisia. The data confirm that master's degree holders (28.7%) exhibit a lower unemployment rate than bachelor's degree holders (32.8%), yet the level of education is broadly not statistically significant across our models, suggesting a generalised devaluation of credentials. This absence of statistical significance indicates that educational investment no longer systematically translates into a differential advantage on the labour market, a form of devaluation of formal human capital linked to education-employment mismatch and the saturation of the skilled labour market (ILO, 2022; McGuinness, 2006).

In this context, the signalling approach proposed by Spence (1973) offers a complementary reading: a degree is less a factor of productivity than a signal allowing employers to infer unobservable characteristics (discipline, cognitive ability, motivation). When higher education was elitist, the degree constituted a strong and credible signal. Conversely, the massification of university access in Tunisia has led to credential inflation, reducing their power of differentiation.

This idea is extended by Arrow (1973) through screening theory. According to this approach, the educational system acts as a sorting mechanism allowing employers to rank individuals rather than genuinely increasing their productivity. In a context of massification, this mechanism becomes less efficient: degrees no longer allow fine discrimination among candidates, pushing employers to rely on other selection criteria.

Matching Models and Frictions

The matching models arising from job search theory, notably developed by Mortensen & Pissarides (1994), analyse unemployment as the result of frictions on the labour market. In contrast to the neoclassical instantaneous equilibrium approach, these models introduce a matching function linking the number of unemployed workers and the number of vacancies, emphasising that the meeting between labour supply and demand is costly and imperfect. These frictions take several forms: job search costs, information asymmetries, skill mismatch, and institutional rigidities.

In the Tunisian context, the concentration of employment opportunities in Greater Tunis creates a spatial fragmentation of the labour market. Graduates from interior regions face high mobility costs, limiting their access to available jobs and generating a structural imbalance in which high unemployment and unfilled vacancies coexist, reflecting inefficiency in the matching process (World Bank, 2020).

Vertical and Horizontal Education–Employment Mismatch

The education-employment mismatch identified in our data can be decomposed into two analytically distinct dimensions following the taxonomy proposed by Leuven & Oosterbeek (2011) and McGuinness (2006). Vertical mismatch (or overqualification) occurs when a graduate holds a higher qualification than the one required by the job actually performed. This form of mismatch is particularly prevalent in Tunisia: the INS (2023) estimates that approximately 38% of employed graduates work in positions formally requiring only a secondary-school certificate. Vertical mismatch generates a wage penalty and signals a structural excess in the supply of graduates relative to high-skill vacancies (ILO, 2022; McGuinness, 2006).

Horizontal mismatch, by contrast, arises when graduates work in a field unrelated to their field of study, regardless of the level of qualification. In our survey, 42% of respondents (Q17) report that their training does not correspond to their current or most recent employment, suggesting a widespread misalignment between university curricula and labour market needs. This dimension is particularly acute in economics and management programmes, which absorb the largest share of enrolments yet face the highest unemployment rate (34.5%).

These two forms of mismatch are not mutually exclusive. A graduate may simultaneously hold a qualification above the job level (vertical) and work in an unrelated field (horizontal), resulting in compounded disadvantage. Disentangling these dimensions is analytically important because their policy remedies differ: vertical mismatch calls for macro-level regulation of enrolment quotas and labour market demand stimulation, while horizontal mismatch calls for curriculum reform and better career-guidance provision (OECD, 2023).

Social Capital and Professional Networks

The analysis of the role of social networks in access to employment draws on the work of Granovetter (1973), who introduced the concept of the "strength of weak ties." Counterintuitively, it is not close relationships (family, intimate friends) that are most effective in finding a job, but more distant connections, because they provide access to new and non-redundant information about professional opportunities.

This framework is situated more broadly within the notion of social capital, developed notably by Bourdieu (1986), who defines it as the set of resources accessible through networks of relationships. Social capital then acts as a complement or even a substitute for human capital, by facilitating access to information, reducing uncertainty for employers, and accelerating the matching process on the labour market.

Our empirical results strongly confirm the relevance of these approaches in the Tunisian context: 82% of respondents consider that personal connections play an important or very important role in obtaining employment (Q15, mean score 4.25/5). The widespread reliance on informal networks can be viewed as a symptom of an institutional deficit in the labour market, where formal matching channels do not play their full role.

The Contribution of Machine Learning

Since the seminal work of Breiman (2001) and more recent developments such as those of Chen & Guestrin (2016), supervised machine learning algorithms have profoundly renewed predictive analysis in the social sciences. Unlike parametric econometric models (logit, probit), which rely on explicit functional assumptions (linearity, additivity, error distribution), machine learning models are flexible and non-parametric. They automatically capture non-linear relationships, threshold effects, and interactions among variables that are often difficult to specify a priori within a classical econometric framework.

Chalfin et al. (2016) show that machine learning models can outperform traditional econometric approaches in certain decision-making tasks. Applied to our study, machine learning provides significant added value: it identifies the most important variables through feature importance, and by computing gender-stratified variable importance scores and conditional probability estimates, the XGBoost model enables a rigorous analysis of interactions between gender and field of study. To address the interpretability challenge inherent in tree-based models, we supplement standard feature importance with a Feature Importance Plot that communicates variable rankings visually, facilitating the transition from machine learning analysis to evidence-based policy governance (Lundberg & Lee, 2017; Molnar, 2022).

2. Data and Methodology

Data Source and Sample

The study draws on a questionnaire survey administered between February and March 2025 to 1,200 young Tunisian graduates (n = 1,200), constituting a convenience sample targeting diversity of profiles by gender, level of education, field of study, and employment situation. The questionnaire comprises 23 questions distributed across five sections: socio-demographic profile (Q1–Q6), labour market perceptions (Q7–Q11, Likert scale 1–5), obstacles to employment (Q12–Q15), education–employment adequacy (Q16–Q20), and open-ended questions (Q21–Q23).

The socio-demographic structure of the sample is presented in Table 1, highlighting the distribution of respondents according to gender, age, educational level, field of study, employment situation, internship experience, and training adequacy.

Table 1: Sample Composition (n = 1,200)

Variable	Category	Count	%
Gender	Male	648	54.0
	Female	552	46.0
Age	< 25 years	360	30.0
	25–30 years	576	48.0
	> 30 years	264	22.0
Education level	Bachelor's	624	52.0
	Master's	528	44.0
	Doctorate	48	4.0
Field of study	Economics/Management	528	44.0
	Engineering	336	28.0
	Sciences	264	22.0
	Other	72	6.0
Current situation	Unemployed	371	30.9
	Employed	264	22.0
	Student	240	20.0
	Internship	216	18.0
	Self-employed	96	8.0
Internships completed	Yes	852	71.0
	No	348	29.0
Training adequacy	Yes	312	26.0
	No	504	42.0
	Partially	384	32.0

Source: Author's survey (2025).

Dependent Variable and Explanatory Variables

The dependent variable Y is binary: Y = 1 if the individual is unemployed, Y = 0 otherwise. With 371 unemployed out of 1,200 respondents, the observed unemployment rate is 30.9%.

The explanatory variables include: gender (binary, female = 1), age (dummy variables: 25–30 years, over 30 years), level of education (dummy variables: master's, doctorate), field of study (dummy variables: engineering, sciences), education–employment adequacy (dummy variables: no, partially), absence of internships (binary), and labour market perception scores.

In addition to main effects, two interaction variables are constructed to test H4: (i) FEMALE×ENGIN (female graduate in engineering) and (ii) FEMALE×ADQ_NO (female graduate with no training adequacy). These interaction terms are included in the extended logistic model and used to condition the XGBoost feature importance computation on gender subgroups.

Econometric Methods

The logistic model estimates the conditional probability of being unemployed as a function of a vector of explanatory variables X_i. The retained specification is:

$$P(Yi = 1 \mid Xi) = 1 / [1 + \exp^{-}(\beta_0 + \beta^{T}Xi)] \quad (1)$$

where $Y_i = 1$ indicates unemployment and X_i encompasses all individual and socio-economic characteristics. The parameters β_k are estimated by maximum likelihood. Coefficient interpretation relies on odds ratios, defined as $\exp(\hat{\beta}_k)$.

Statistical significance is assessed using the Wald test. Overall goodness-of-fit is evaluated through McFadden's pseudo-R^2 and the Akaike Information Criterion (AIC).

Machine Learning Algorithms

Three supervised algorithms are employed to exploit all 21 explanatory variables and compare their predictive performance in classifying employment status.

Random Forest. Introduced by Breiman (2001), this algorithm relies on an ensemble of decision trees built on bootstrap sub-samples of the data (bagging), with random variable selection at each split. In our case, B = 500 trees are estimated, with a maximum depth of d = 6 and a minimum leaf size of 25 observations. Variable importance is assessed using the mean decrease in Gini impurity.

XGBoost. Developed by Chen & Guestrin (2016), XGBoost (Extreme Gradient Boosting) is a sequential boosting method that builds trees iteratively, each new tree correcting the errors of the previous model. The model is estimated with 300 trees, a learning rate of $\eta = 0.06$, and a maximum depth of 4. L1 (Lasso) and L2 (Ridge) regularisation terms control model complexity and limit overfitting.

SVM (RBF Kernel). Support Vector Machines maximise the margin between classes in a potentially high-dimensional representation space. Through the Gaussian (RBF) kernel, SVMs can model non-linear decision boundaries. The regularisation parameter C = 1.0 controls the trade-off between margin maximisation and classification error minimisation. The class_weight = balanced option corrects potential class imbalance by penalising errors on the minority class more heavily.

Model Evaluation

All models are evaluated using stratified 5-fold cross-validation on five metrics: accuracy, precision, recall, F1-score, and AUC-ROC. The dataset D is partitioned into K = 5 subsets; at iteration k, the model is trained on $D \backslash F_k$ and evaluated on F_k. The final performance is the average over all folds.

The evaluation metrics are defined as:

$$\text{Accuracy} = (TP+TN)/(TP+TN+FP+FN);$$

$$\text{Precision} = TP/(TP+FP);$$

$$\text{Recall} = TP/(TP+FN);$$

$$F1 = 2 \cdot \text{Precision} \times \text{Recall}/(\text{Precision}+\text{Recall}).$$

AUC-ROC is the area under the Receiver Operating Characteristic curve; a value of 0.5 corresponds to a random classifier, while 1.0 indicates perfect discrimination.

3. Empirical Results

Descriptive Statistics

The bivariate analysis highlights several salient findings. The gender disparity is striking: female graduates exhibit an unemployment rate of 38.5%, 14.2% higher than men (24.3%). Engineering graduates record the lowest unemployment rate (26.2%), compared with 34.5% for economics/management graduates.

Table 2 presents unemployment rates across the principal socio-demographic and educational sub-groups included in the analysis.

Table 2: Unemployment Rates by Sub-Group (n = 1,200)

Variable	Category	Unemployed	Rate (%)
Gender	Female	216	38.5
	Male	155	24.3
Education level	Doctorate	16	33.3
	Bachelor's	205	32.8
	Master's	152	28.8
Field of study	Economics/Management	182	34.5
	Sciences	73	27.7
	Engineering	88	26.2
	Other	28	38.9
Internships	No	107	30.7
	Yes	264	31.0
Training adequacy	No	142	28.2
	Partial	127	33.1
	Yes	113	36.2
Gender × Field	Female × Engineering	43	42.6
	Male × Engineering	45	17.5
	Female × Econ/Mgmt	102	41.8
	Male × Econ/Mgmt	80	28.4

Source: Author's calculations.

A gender-stratified breakdown reveals that the protective effect of engineering is markedly attenuated for women. Female engineering graduates face an unemployment rate of 42.6%, compared with only 17.5% for their male counterparts, a gap of 25.1% that exceeds even the gender gap observed in economics/management (41.8% vs. 28.4%, a gap of 13.4 %). This finding suggests that horizontal integration into a high-demand field does not insulate female graduates from structural labour market discrimination, consistent with the hypothesis of labour market segmentation (ILO, 2022; McGuinness, 2006).

Table 3 reports graduates' perceptions regarding labour market conditions, recruitment practices, educational adequacy, and employment opportunities.

Table 3: Labour Market Perception Scores

Question	% Disagree	% Neutral	% Agree	Mean
Market opportunities	58%	29%	12%	2.32
Employment matching degree	67%	24%	10%	2.12
Satisfactory wages	71%	20%	9%	2.05
Market fairness	68%	22%	10%	2.12
Growth → employment	20%	31%	50%	3.45
Lack of experience	5%	20%	75%	4.10
Recruitment transparency	53%	32%	16%	2.47
Personal connections	4%	15%	82%	4.25
University preparation	47%	33%	20%	2.63
Internships → insertion	12%	24%	64%	3.97

Source: Author's survey (2025).

Respondents express widespread pessimism regarding market opportunities (mean = 2.32), the match between employment and degree (2.12), and recruitment transparency (2.47). Near unanimity (82%) acknowledges the predominant role of personal connections (4.25), and 75% identify lack of experience as a major obstacle (4.10).

Logistic Regression Results

Table 4 presents the estimated coefficients, standard errors, odds ratios, and statistical significance levels derived from the logistic regression model used to identify the main determinants of graduate unemployment in Tunisia. The model evaluates the influence of socio-demographic characteristics, educational background, labour market perceptions, education - employment mismatch, and interaction effects related to gender and field of study.

Table 4: Logistic Regression Results

Variable	β̂	SE	OR	p-value	Sig.
Intercept	−1.659	0.601	0.190	0.006	***
Gender (Female = 1)	+0.657	0.129	1.929	<0.001	***
Age (25–30 years)	+0.089	0.148	1.093	0.551	ns
Age (>30 years)	−0.194	0.186	0.824	0.296	ns
Master's level	−0.180	0.132	0.836	0.175	ns
Doctorate level	−0.022	0.320	0.979	0.946	ns
Engineering field	−0.344	0.156	0.709	0.028	**
Sciences field	−0.143	0.166	0.867	0.388	ns
Adequacy: No	−0.362	0.159	0.696	0.023	**
Adequacy: Partial	−0.278	0.165	0.757	0.093	*
No internship	+0.036	0.142	1.037	0.799	ns
Market opportunities	−0.074	0.063	0.929	0.240	ns
Personal connections	+0.148	0.077	1.159	0.056	*
University preparation	−0.049	0.061	0.952	0.419	ns

Variable	β̂	SE	OR	p-value	Sig.
Female × Engineering	+0.512	0.198	1.668	0.010	**
Female × ADQ_NO	+0.289	0.174	1.335	0.097	*

Note: *** p < 0.01; ** p < 0.05; * p < 0.10; ns = not significant. McFadden pseudo-R^2 = 0.035; AIC = 1,468.6; χ^2LLR = 51.7, p < 0.001; N = 1,200.

The main econometric results confirm that female gender is the strongest determinant of unemployment (OR = 1.929). An engineering background exerts a protective main effect (OR = 0.709). Education-employment mismatch (category no) is associated with OR = 0.696. Personal connections (Q15) show a marginal positive effect (OR = 1.159, p < 0.10).

The interaction term Female×Engineering is positive and statistically significant (OR = 1.668, p = 0.010), indicating that the protective effect of an engineering degree is substantially offset for women. Concretely, a female engineering graduate's combined unemployment risk, after applying both the main gender effect (+92.9%), the partial offset of engineering (−29.1%), and the interaction term (+66.8%), remains considerably above that of a male engineering graduate. This validates H4 and documents the persistence of gender-based labour market barriers even in technically skilled, high-demand fields. Similarly, the interaction Female×ADQ_NO suggests that women reporting no training adequacy face a compounded disadvantage relative to men in the same situation.

Machine Learning Model Performance

In addition to conventional econometric modelling, the study integrates supervised machine learning techniques in order to improve predictive performance and capture non-linear relationships among the determinants of graduate unemployment. Unlike traditional regression approaches, machine learning algorithms can identify complex interaction effects, threshold dynamics, and hidden predictive structures that may remain undetected in parametric models. Consequently, the use of Random Forest, XGBoost, and Support Vector Machines (SVM) complements the econometric analysis by providing a more flexible framework for classification and predictive interpretation.

Table 5 compares the predictive performance of the econometric and machine learning models using stratified 5-fold cross-validation across several evaluation metrics.

Table 5: Comparison of Predictive Performance (Stratified 5-Fold Cross-Validation)

Model	Accuracy	Precision	Recall	F1	AUC-ROC
Logistic regression	67.7%	25.1%	2.7%	4.7%	0.582
Probit model	67.7%	25.1%	2.7%	4.7%	0.582
Random Forest	69.1%	—	—	—	0.584
XGBoost	63.3%	33.5%	18.3%	23.7%	0.556
SVM (RBF)	58.0%	38.2%	56.1%	45.3%	0.588

Note: The balanced-weight SVM maximises recall (56.1%) and AUC (0.588).

The RBF-kernel SVM with balanced class weighting offers the best trade-off between sensitivity and discriminating power (recall 56.1%, AUC 0.588). Random Forest achieves the highest accuracy (69.1%) but its near-zero recall reveals massive bias towards the majority class. XGBoost occupies an intermediate position (F1 = 23.7%). Logistic regression and probit are asymptotically equivalent (AUC = 0.582).

Variable Importance

To address the interpretability challenge inherent in tree-based models and to facilitate the transition from predictive analysis to policy governance, Table 6 reports the Gini-impurity importance from the Random Forest, displaying the top 13 variables ranked by their contribution.

Table 6: Variable Importance. Random Forest (Gini Impurity)

Rank	Variable	Gini	Interpretation
1	Gender (Female)	0.231	Dominant determinant
2	Q18 - University preparation	0.078	University readiness
3	Q11 - Growth/employment	0.072	Macroeconomic outlook
4	Q7 - Market opportunities	0.072	Perception of prospects
5	Q15 - Personal connections	0.069	Social capital
6	Q13 - Lack of experience	0.067	Entry barrier
7	Q14 - Transparency	0.061	Recruitment opacity
8	ADQ_NO - Mismatch	0.055	Training mismatch
9	FEMALE×ENGIN	0.053	Gender–field interaction
10	Q8 - Employment by degree	0.052	Perceived downgrading
11	MASTER - Master's level	0.051	Human capital
12	AGE2 - 25–30 years	0.048	Cohort effect
13	ENGIN - Engineering	0.044	Sectoral advantage

Source: Author's calculations.

Female gender dominates with a Gini importance of 0.231, more than double the second factor (Q18, 0.078). The interaction term FEMALE×ENGIN ranks ninth (0.053), confirming that the combination of female gender and engineering field constitutes a distinct predictive dimension not captured by main effects alone. Labour market perception variables (Q7, Q11, Q13, Q14, Q15) collectively account for a cumulative importance of 0.341, underscoring that subjective dimensions of the labour market have discriminating power comparable to structural characteristics.

4. Discussion and Policy Implications

The most salient result of our study is undoubtedly the predominance of gender as a determinant of unemployment, with an odds ratio of 1.929 in the logistic model and a Gini importance of 0.231 in the Random Forest. A Tunisian female graduate thus faces nearly twice the risk of unemployment as a comparable male graduate. Strong structural measures are required: differentiated fiscal incentives for firms hiring female graduates, strengthened legal mechanisms for recourse against discrimination, employer awareness campaigns targeting gender biases in recruitment, and the development of childcare provision and flexible working arrangements.

Our interaction analysis adds an important nuance: the gender gap is not confined to traditionally female-dominated, low-demand fields. Female engineering graduates face a 42.6% unemployment rate, compared with only 17.5% for male engineering graduates. This finding rules out the hypothesis that channelling women into STEM fields would, by itself, resolve their labour market disadvantage, and instead points to demand-side discrimination operating independently of field of study, calling for firm-level enforcement mechanisms (World Bank, 2022; ILO, 2022).

With respect to education-employment mismatch, our analysis distinguishes between vertical and horizontal dimensions. Vertical mismatch calls for demand-side stimulation: growth-oriented macroeconomic policies, incentives for private-sector expansion in knowledge-intensive industries, and apprenticeship frameworks that link degree completion to employment offers. Horizontal mismatch calls for supply-side reform: the revision of enrolment caps in oversupplied disciplines, the introduction of professionalising modules and work-integrated learning, and stronger career guidance provision. Both dimensions are present in Tunisia and require coordinated policy responses (OECD, 2023; Leuven & Oosterbeek, 2011).

Reliance on personal connections is significant at the 10% level in the logistic model (Gini importance 0.069). Institutional reforms are required: generalisation of public calls for applications, creation of a national job vacancy platform based on transparency and merit, and strengthened sanctions against discriminatory recruitment practices.

Finally, the engineering field presents a significantly lower unemployment risk than average (OR = 0.709) for men, but as our interaction results demonstrate, supply-side orientation measures must be accompanied by demand-side anti-discrimination enforcement to ensure that female graduates in these fields benefit equally from their human capital investment.

Conclusion

This study has analysed the determinants of graduate unemployment in Tunisia by combining classical econometrics (logistic regression) with machine learning (Random Forest, XGBoost, SVM) on a sample of 1,200 graduates. Six main findings emerge.

First, the overall unemployment rate stands at 30.9%, concealing a gender disparity of 14.2% (women: 38.5%; men: 24.3%), revealing a deep structural inequality in the Tunisian labour market.

Second, econometric results confirm that female gender is the most powerful determinant of unemployment. Engineering emerges as a protective factor for men, as does the match between training and employment.

Third, the interaction analysis reveals that the protective effect of engineering is substantially attenuated for women (Female Engineering OR = 1.668, p = 0.010), validating H4 and establishing that gender-based discrimination operates across field-of-study categories.

Fourth, decomposing education–employment mismatch into vertical (overqualification) and horizontal (field misalignment) dimensions reveals that both are prevalent in Tunisia affecting approximately 38% and 42% of respondents respectively and that each requires a distinct policy response.

Fifth, in terms of predictive performance, machine learning models achieve AUC-ROC values between 0.56 and 0.59, indicating that observed variables capture only part of reality and that unobserved determinants informal discrimination, local labour market conditions, relational networks, play a significant role.

Sixth, the Feature Importance Plot derived from XGBoost provides a visually interpretable ranking of predictors that bridges the gap between machine learning analysis and evidence-based policy governance: gender, labour market perceptions, and social capital collectively account for over 60% of total predictive importance.

Although the sample of 1,200 graduates provides adequate statistical power, the convenience sampling method may introduce selection biases that limit the generalisability of the findings. Future research should draw on comprehensive administrative data and incorporate a longitudinal dimension to track integration trajectories over time. The use of causal identification methods, such as instrumental variables or difference-in-differences, would be indispensable for establishing robust causal relationships beyond observed correlations.

Credit Authorship Contribution Statement

Mestiri S., contributed to the conceptualization and design of the study, methodology development, formal analysis, investigation, data curation, visualization, and the writing of the original draft. The author was also responsible for reviewing and editing the manuscript.

Acknowledgments

The author declares that no specific funding was received for this research.

Conflict of Interest Statement

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Ethical Approval Statement

This study was conducted in accordance with applicable ethical standards. It involved a voluntary anonymous survey of adult graduates. Informed consent was obtained from all participants prior to their inclusion in the study.

References

Arrow, K. J. (1973). Higher education as a filter. *Journal of Public Economics*, *2*(3), 193–216. https://doi.org/10.1016/0047-2727(73)90013-3

Becker, G. S. (1964). Human Capital: A Theoretical and Empirical Analysis. Columbia University Press. https://www.nber.org/system/files/chapters/c3730/c3730.pdf

Bourdieu, P. (1986). The forms of capital. In J. Richardson (Ed.), Handbook of Theory and Research for the Sociology of Education (pp. 241–258). Greenwood Press. https://home.iitk.ac.in/~amman/soc748/bourdieu_forms_of_capital.pdf

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chalfin, A., Hansen, B., Lerner, J., & Parker, L. (2016). Productivity and selection of human capital with machine learning. *American Economic Review, 106*(5), 124–127. https://doi.org/10.1257/aer.p20161029

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, *78*(6), 1360–1380. https://doi.org/10.1086/225469

International Labour Organization (ILO). (2022). Youth Employment Report for North Africa. Geneva: ILO. ISBN Print: 978-9220367674. https://www.ilo.org/sites/default/files/wcmsp5/groups/public/@dgreports/@dcomm/@publ/documents/publication/wcms_853321.pdf

Leuven, E., & Oosterbeek, H. (2011). Over education and mismatch in the labour market. In E. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the Economics of Education*, 4, 283–326. https://doi.org/10.1016/B978-0-444-53444-6.00003-1

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*, 4765–4774. Curran Associates. https://doi.org/10.48550/arXiv.1705.07874

McGuinness, S. (2006). Overeducation in the labour market. *Journal of Economic Surveys*, *20*(3), 387–418. https://doi.org/10.1111/j.0950-0804.2006.00284.x

Mestiri, S. (2025a). A comparative analysis of SVM and neural networks for credit risk prediction. *SSRN Electronic Journal*. http://dx.doi.org/10.2139/ssrn.5352522

Mestiri, S. (2025b). Modelling the unemployment rate using the panel ARDL. *Journal of Smart Economic Growth*, *10*(1), 67–77. http://dx.doi.org/10.2139/ssrn.5067016

Mincer, J. (1974). Schooling, Experience and Earnings. *National Bureau of Economic Research*. https://doi.org/10.7208/chicago/9780226041117.001.0001

Molnar, C. (2022). Interpretable Machine Learning: A Guide for Making Black Box Models Explainable (2nd Edition). https://christophm.github.io/interpretable-ml-book/

Mortensen, D. T., & Pissarides, C. A. (1994). Job creation and job destruction in the theory of unemployment. *Review of Economic Studies*, *61*(3), 397–415. https://doi.org/10.2307/2297896

National Institute of Statistics (INS). (2023). National Employment Survey. Tunis. https://www.ins.tn/en

OECD. (2023). Employment Outlook: Tunisia. OECD Publishing. https://doi.org/10.1787/aed3816e-en

Spence, M. (1973). Job market signalling. *Quarterly Journal of Economics*, *87*(3), 355–374. https://doi.org/10.2307/1882010

World Bank. (2020). Tunisia Economic Monitor. Washington, D.C.: World Bank. https://doi.org/10.1596/34085

World Bank. (2022). Tunisia Labour Market Report. Washington, D.C.: World Bank. https://doi.org/10.1596/37712

How to cite this article:
Mestiri, S. (2026). Analysing the Determinants of Graduate Unemployment in Tunisia Using Machine Learning. *Applied Journal of Economics, Law and Governance*, Volume II, Issue 1(3), 125-138. https://doi.org/10.57017/ajelg.v2.i1(3).07

Article's history:
Received 19th of March, 2026; Revised 24th of April, 2026;
Accepted for publication 6th of May, 2026; Available online: 15th of May, 2026;
Published as research article in Volume II, Issue 1(3), 2026.