Behavioral Predictors of Academic Performance: Interpretable Neural Network Models of GPA Using Large-Scale Survey Data

J. Fernando Plascencia Guzmán* Willy Walter Cortez Yactayo**

*Postdoctoral Researcher Secihti, CUCEA, CEPRODE, University of Guadalajara
Corresponding author: juan.plascencia@academicos.udg.mx

**Professor-Researcher SNI III, CUCEA, Director of CEPRODE, University of Guadalajara

Abstract

The present reseach aims to identify the determinants of academic achievement among sixth-grade students (10-15 years old) in Mexico using machine learning techniques applied to large-scale survey data. We trained neural network to predict academic performance defined as the average test scores of Mathematics and Language & Communication (L&C) from the Achievement Assessment Related to the National Education System (ELSEN). The models incorporate multidimensional explanatory variables from two national cohorts (2014–2015 and 2017–2018; N = 104,204 and N = 104,973, respectively). In addition, we use a local interpretable model-agnostic explanation (LIME) to capture feature importance that allow the identification of the most influential predictors. Our findings show that academic achievement is shaped by a set of behavioral, environmental, and socioeconomic factors rather than cognitive ability alone. Factors such as home learning environments, school climate, teacher–student relationships, socioeconomic conditions, access to cultural enrichment, and early childhood education appears as determinats of academic achievement. These results highlight the importance of understanding educational outcomes as the product of layered interactions across family, school, and broader social contexts and from a policy perspective the need for integrated, cross-sectorial strategies.

Keywords: Education, Machine Learning.

1 Introduction

Understanding the determinants of academic performance remains a central question in education research and policy [Coleman et al., 1966, OECD, 2013]. Cognitive ability (CA) and socioeconomic status (SES) have traditionally been seen as primary influences [Sirin, 2005, Duncan and Murnane, 2011], other research pointed to explore behavioral, motivational, and contextual factors [Zimmerman, 2002, Duckworth et al., 2007]. There are aggregated measures of SES and how SES disparities in academic achievement can be explained by differences in the quality of education that a child received. Some researchers use SES and social class interchangeably (wealth, power, and social status) [Mueller and Parcel, 1981], others consider SES as a tripartite nature (parental income, parental education, and parental occupation) [Duncan et al., 1972]. The multifactorial causes of academic achievement ranges from demographic and household factors such as race, gender, ethnicity, parental education, and socioeconomic status that are strongly linked to educational outcomes [Entwisle et al., 2005, Rumbaut, 2005]. These studies have shown a medium to strong achievement relation, but is contingent upon factors such as the range of socioeconomic variables used and the type of these variables with the achievement measure, school level, minority status, and school location.

Another determinant is the school's climate and institutional priorities, which influence academic achievement, often privileging high achievers and creating inequities [Attewell, 2001, Wang and Degol, 2016, Bouchouna, 2024]. Other studies have found that peer influence can affect performance, and some others have tested the effects of parental involvement [Mounts and Steinberg, 1995, Fite et al., 2014, Ahmetoglu et al., 2020]. In this area, some researchers have found that adolescents' school achievement is affected by peer behavior both positively and negatively. High achievement peers have positive effects on adolescents' satisfaction with school, educational expectation, grades, and test scores [Epstein and N., 1983],contrasting with peer drug use, alcohol, antisocial behavior and poor academic achievement [Mounts and Steinberg, 1995].On the other hand, parenting styles relate to the way adolescents develop attachments to their peers and to academic self-efficacy. For example, researchers have found that a parental permissive style is an important positive predictor of aggressive behavior and a negative predictor of attachment to their peers [Llorca et al., 2017]. In addition, the other issue is parental involvement on children´s academic performance and peer victimization. One example is that ostracizing from peers can over relational or cyber and is associated with poor adjustment outcomes. High levels of peer victimization are associated with lower academic performance, however, effect sizes are inconsistent across studies. Some researchers have evaluated over victimization, relational victimization, parental involvement, and academic performance. ¨ High levels of peer victimization, particularly relational victimization, were associated with lower levels of academic performance at both high and low level of parental involvement. However, the highest levels of academic performance were evident when parental involvement was high, and levels of relational victimization were low, and the lowest levels of academic performance occurred when parental involvement was low, and levels of relational victimization were high¨ [Fite et al., 2014] .

On the other hand, from the perspective of behavioral frame, psychological traits like self-efficacy, discipline, and motivation also predict academic success[Komarraju et al., 2011, Komarraju et al., 2013, Caraway et al., 2003, Duckworth and Seligman, 2005, Duckworth et al., 2007, Wolters, 1999, Tucker-Drob and Harden, 2012, Tucker-Drob and Bates, 2015, West et al., 2016]. There are several studies that incorporate the role of parental involvement in homework (autonomy support, control interference, cognitive engagement), student homework behavior, and student academic achievement, and how these relationship vary depending on school level. [Gonida and Cortina, 2014, Nuñez et al., 2015, Wang and Sheikh-Khalil, 2014]. In contrast, autonomy support and avoidance of interference have been associated with positive outcomes, while parental control and direct aid have been shown to be detrimental. However, achievement goal theory pointed out why students engage in academic behaviors. Goal orientations such as mastery and performance are associated with different cognitive, affective and behavioral patterns of learning. As a social cognitive theory, achievement goal theory asserts that goal orientations become part of a student's life within the context of school and family [Gonida and Cortina, 2014, Maehr, 2001]

Furthermore, in parenting style research has found that it has an important influence on academic achievement, not only by family´s social background but also by a combination of parents' current class location and their own family backgrounds. [Roksa and Potter, 2011] and peer relationships. Parental styles differ in terms of the levels of parental sensitivity (warmth and affection), and parental control (promoting children's autonomy), and both factors are associated with child development. The theory of parenting style is focused on the control of parents over their children or parental demandingness and on the level to which parents respond to child´s needs or parental responsiveness. There are three different parenting styles that have been identified that emerge as the combination of these factors: authoritative (moderate demandingness and moderate responsiveness), permissive (low demandingness and high responsiveness) and authoritarian (high demandingness, and low responsiveness) [Baumrid, 1966]. The impact of parents with higher education is more likely to support learning through strategic behaviors [Vellymalay, 2012, Baker and Stevenson, 1986]. The findings on the relationship between parental involvement, parental support, and family education on pupil achievement and adjustment in schools shows that Parental involvement takes various forms, such as creating a supportive home environment, encouraging learning, maintaining communication with schools, attending events, contributing to school activities, and participating in school governance. The level and type of involvement are influenced by factors like social class, the mother's education and mental health, financial hardship, single parenthood, and, to a lesser extent, ethnicity. Children's academic success often leads to greater parental involvement, and supportive parenting at home has a strong positive effect on achievement, more so than the quality of the school itself, especially in primary education. However, disparities in parental involvement often stem from socioeconomic challenges, health issues, and parents' confidence or understanding of their role. Some parents may feel discouraged by negative interactions with schools or staff. [Desforges, 2003].

Although, family structure has an important role: children in single-parent homes or with incarcerated fathers tend to have lower GPAs [Astone and McLanahan, 1991, Rodgers and Rose, 2001, Sun and Li, 2011, Foster and Hagan, 2009, Cavanagh et al., 2006]. In the U.S., "differences in ethnicity, gender, and generation shape the socioeconomic trajectories of young adults, particularly in education, incarceration, and early childbearing. Among Latin American and Asian-origin youth, there is overall educational improvement from immigrant to U.S.-born generations . However, significant variation exists by ethnic background and gender, with Asian groups and females generally achieving higher educational levels. These patterns highlight persistent inequalities in social and economic outcomes. There is a strong association between low education levels and high rates of incarceration among men and early childbearing among women, which are linked to reduced occupational and economic success. Additionally, U.S.-born women show lower childbearing rates than immigrant women"[Rumbaut, 2005]. Economic resources and parental behavior explanations for family structure has an important effect on children: "single-mother families and never-married single mothers are similar in the extent to which economic disadvantage accounts for poorer child outcomes. In addition, children living with their stepfathers or mothers' cohabiting partners and lower levels of maternal as well as paternal support in these families, compared to original two-parent families, appear to contribute to problem behaviors and less engaging temperaments. Although single mothers who have never married and those who were previously married share many similarities, there are notable differences. The most significant is that never-married mothers tend to have very low education levels (on average, less than a high school diploma), higher rates of unemployment, and fewer hours worked. While unemployment leads to greater maternal support and involvement, the lack of education has negative impacts on their children".[Thomson et al., 1994]

Finally, participation in arts, music, and cultural programs contributes to academic engagement, academic aspitations and outcomes, however, it depends on the context in which this participation occur (i.e. school , home or community). [Martin et al., 2013].

Considering all these contextual factors grounded mostly in the USA case, the aim of this research is to identify the determinants of academic achievement in Mexico, narrowing our analysis for elementary 6 grade students. The Mexican National Education System differs significantly from those of the United States and Europe. Educational researchers have long emphasized that factors such as a student's socioeconomic status (SES) influence both the quality of schools attended and the neighborhoods in which students reside [Caldas and Bankston, 1997, Brooks-Gunn et al., 1997]. In the Mexican context, the more influencial SES are household income variable, educational level, social gap,urban or rural residency. To illustrate a basic income differences, data from the National Institute of Statistics and Geography (INEGI) indicate that being considered middle class entails an average monthly household income of approximately 22,297 pesos (1,114 USD). This figure varies by region, with urban households averaging around 23,451 pesos (1,172 USD) and rural households approximately 18,569 pesos (928.45 USD). The National Institute of Statistics and Geography 2025 consider a lower-class households those who earn an average of 11,343 pesos per month (567.15 USD), a middle-class households around 22,297 pesos (1,114 USD), and upper-class households 77,975 pesos or more (3,898+ USD).

In contrast to other educational systems, Mexican families do not necessarily relocate to high-SES neighborhoods to improve their children's access to quality education. Instead, when financial conditions allow, private schooling is frequently chosen. School choice is often influenced more by geographic proximity to the household or parents' workplaces than by neighborhood socioeconomic characteristics. The Mexican case reflects unique patterns in access to educational quality, which have been widely discussed for the Americam case in the literature. These disparities are not primarily driven by race or gender—as is often the case in other countries—but are instead associated with factors such as geographical area (urban or rural), ethnicity (particularly membership in a Mexican indigenous group), and family income.

The analyzes of the determinants of academic achievement of previous research in this area uses both parametric and non-parametric methods. However, these methods are limited to explore complex dimension of factors that might explain academic performance. In this paper, we propose a methodological approach grounded in machine learning, using neural network models to predict Grade Point Average (GPA) among sixth-grade students in Mexican elementary schools. To enhance the interpretability of the model's outputs, we apply Local Interpretable Model-agnostic Explanations (LIME) to identify the features most frequently and consistently associated with predicted GPA [Ribeiro et al., 2016, Molnar, 2022]. Rather than testing predefined hypotheses, this study aims to uncover insights into the behavioral, social, and contextual factors that influence academic performance, guided by established theoretical perspectives [Bandura, 1986, Kahneman, 2011]

2 Data and Methods

2.1 Description of Variables and Sample Information

The dataset used in this study was provided by the former Instituto Nacional para la Evaluación de la Educación (INEE), the Mexican national authority that used to be responsible to the application of the Achievement Assessments related to the National Education System (ELSEN). The dataset is composed of several integrated components which can be classified as follows: (1) a background module containing demographic and contextual variables; (2) a survey module addressing dimensions of family life and child development; (3) a test module providing item-level data on academic assessments; and (4) an evaluation module summarizing test performance. The survey module captures a broad set of constructs, including household demographics, parental relationships, parent–child interactions, indicators of child well-being and behavior, health status, socio-economic conditions (e.g., parental employment, income, housing quality, and access to resources), and features of the school environment.

The test module includes item-level responses and performance metrics in Mathematics and Language and Communication for the 2015 cohort, and additional subjects in the 2018 administration. The databases were intended for the evaluation of the National Education System (ELSEN). Therefore, were used estimates of parameters that characterize the various subpopulations of the country in these assessments. To achieve this, the scales were constructed using complex statistical methods commonly used in assessments of student academic performance around the world and are known as "plausible values." In this case, five plausible values were generated for each student's achievement variable in each test, based on their responses to the achievement items and the contextual questionnaire administered to the sampled students. This reserach focuses on two student cohorts:

- **2015 Cohort:**Consisting of 104,204 sixth-grade students from 3,446 primary schools. Students were assessed in Mathematics and Language and Communication and age 10-15 years old.

INEE states thatf for his cohort the sample was designed to produce results at the national level, by type of school, and by federal entity. The number of students selected in the sample was calculated to have a maximum margin of error of 10% of the standard deviation of the achievement variable of interest (means, percentages, correlations) at the federal entity level. For results at a lower level of disaggregation than the federal entity—such as school types within a federal entity or other subpopulations—the margin of error increases due to the smaller amount of data available for estimating population parameters. For each of the assessments, the mean of the study population was set at 500 units with a standard deviation of 100 units. This allows any difference to be quantified in terms of the standard deviation (effect size). In addition to the databases, the main results of the assessments are included, along with their corresponding standard errors and the amount of information used in the estimation.

- **2018 Cohort:** Comprising 108,083 students from 3,573 schools. Of these, 104,973 students completed assessments in Mathematics and Language and Communication, 93,992 in Civics and Ethics, and 7,901 in Written Expression. While the 2017 survey instrument differs slightly in structure from the 2015 version, it retains thematic continuity in its coverage of key constructs. The students age 10-15 years old.

INEE states that another type of variable included in the data are indices, which are derived from sets of questions from the student questionnaire that summarize information on various contextual aspects. These indices allow for the exploration of the relationship between achievement and students' school and family environments. They are based on population distributions with a mean of 50 units and a standard deviation of 10 units for each index. The data correspond to a sample drawn from the target population. This sample was designed to produce results for the following subpopulations:

For all subjects:

– Students nationwide

– Students by type of school (4 types: General Public, Indigenous, Community-based, and Private)

– Students by level of marginalization

– Students by locality size

– Students by rural-urban population

For Language and Communication, Mathematics, and Civic and Ethical Education:

– Students by federal entity

For Language and Communication and Mathematics:

– Students by federal entity and type of school

– Students by federal entity and level of marginalization

– Students by federal entity and rural-urban population

The sampling design of the survey made by INEE was carried out in two stages using random selections. In the first stage, schools were selected; in the second stage, students were selected. The number of students chosen was planned to achieve a margin of error of 12% of the standard deviation of achievement when estimating the mean at the federal entity level. For results involving aggregations of smaller groups than those defined in the sample, it were recommended to review the size of the associated standard errors before making any interpretations. In general, the estimation results—both in terms of achievement and context—require the use of expansion factors to ensure that the selected students adequately represent the target population and to estimate the standard errors. For additional achievement results, appropriate methodologies for plausible values should be used.

We are interested in identify general determinants of academic achivement. For the aim of this research, we computed a general average of the five plausible values in Mathematics and Language and Communication using item-level performance data from the respective academic years (2014–2015 and 2017–2018). The analytical sample used in this resarch includes students who completed both subject tests (N = 104,204 for 2015; N = 104,973 for 2018). These averaged scores were used as the primary dependent variable in the predictive modeling of academic performance that we called a Grade Point Average (GPA).

2.2 Methods

To investigate the factors influencing academic performance, we designed and estimated a series of feedforward neural network models. The input data used for training corresponds to cohorts from the 2015 and 2017 academic years, respectively. Each dataset was cleaned and transformed separately due to the differences in the survey questions posed to each cohort, although the same cleaning and transformation procedures were applied consistently. There are multiple frameworks available for designing neural networks to predict academic performance [Baashar et al., 2022, Rodríguez-Hernández et al., 2021, Ahmad and Shahzadi, 2018, Cascallar et al., 2015] some of these works have compared traditional parametric approaches and compared the results with neural networks [Hardgrave et al., 1994]; based on the framework suggested by [Lau et al., 2019, Davidson, 2019, Musso et al., 2013, Musso et al., 2020] which objective is similar to our study, we followed them in a similar procedure. However, we extend the previously proposed parameter-combination approach to a different dataset. The codebase implementation in [Davidson, 2019] was no longer compatible with current libraries, so the codebase was updated to modern frameworks to ensure both functionality and reproducibility.

We begin by inspecting the data to identify which columns will be used as input features and to detect missing values. Since our goal is to predict academic performance, we identified a dependent variable derived from the columns of the plausible values for LyC and Math tests. These values represent the potential levels of academic performance in these subjects, and we calculate an average mean of both as a general achievement level. In cases where a student completed only one subject, we used the average of the plausible values for that subject's test. The next step involved categorizing every column as either categorical or continuous, which informed the imputation of missing values. For this purpose, we also implemented a k-nearest neighbor (k-NN) imputation procedure, where the values of the five closest neighbors for each observation were used [Rubinsteyn and Feldman, 2016]. Unlike the approach of [Davidson, 2019] continuous values were imputed using the mean of the neighbors, while categorical values were imputed by temporarily converting them to integers and then mapping the imputed values back to the closest original category after the k-NN procedure (we don not use the mode). It is a common procedure that continuous features were standardized by subtracting the mean and dividing by the standard deviation [LeCun et al., 2015]. All categorical variables, including ordinal types, were transformed using one-hot encoding, resulting in a binary indicator matrix where each column corresponds to a distinct category level. This encoding scheme represents the presence or absence of a specific category using separate binary features (dummy variables). In the resulting feature matrix, each original categorical variable is mapped to one or more derived features, with reference categories omitted to avoid multicollinearity. Additionally, the final training data sets for both years were subdivided into 80% for model fitting and 20% for validation.

In traditional neural networs design, key parameters are weights and biases that are adjusted during the training process improving a model's perfomance. The selection of such paramenters determine how the network learns from data. We designed a network function based on the module Keras [Chollet, 2015] with TensorFlow [Abadi et al., 2015] which architecture was parameterized through a model function, which enables the iteration of defined hyperparameters, in detail we include four activation functions (linear, sigmoid, ReLU, tanh), which has an impact on how the network learms and introduce non-linearity into the network allowing it to learn complex patters; another parameter is the number of hidden layers (we choose a set ranging from zero to three), another hyperparameter is the size of each hidden layer (64, 128, or 256 neurons) [Hornik et al., 1989]. In contrast to other methodological approaches, like in [Davidson, 2019], we incorporate different dropout rates (0.1, 0.2, or 0.3) which will allow to mitigate overfitting and regularization, and batch normalization layers to stabilize and accelerate training dynamics [Srivastava et al., 2014]. There are different optimization algotithms, (they adjust weights and biases during tranning to minimize the difference between predictions and actual target values),we choose the Adam optimization algorithm to minimize mean squared error loss, which is appropriate for our data structre and for a continuous outcome prediction [Kingma P and Lei Ba, 2015]. Furthermore, the training data was conducted with early stopping hyperparameter based on validation loss to prevent overfitting and enhance generalization. In addition, we implemented a 5-fold cross-validation to test results. This technique allow us to provide robust estimates of the model's generalization performance, employing negative mean squared error as the scoring metric, which is suitable for regression tasks.

This systematic exploration via gridsearch in the sickitlearn environment , allow us to identified optimal configurations of activation functions, network depth, layer sizes, and dropout rates. Once the five best-performing models were retrained, the final outcomes of these models, "the keras models" will serve as an input for an agnostic interpretability analysis. The interpretability of these models is addressed through the application of Local Interpretable Model-agnostic Explanations (LIME) [Ribeiro et al., 2016]. That is, the network architecture will provide keras selected models and the LIME algorithm will take these models and provide insights into individual model predictions. The explanation behind this is that an agnostic model approximate the behavior of a complex model in the local vicinity of a specific instance using a simpler, interpretable surrogate model. This method works by perturbing the input data and observing the resulting predictions from the black-box model. A weighted interpretable model—typically linear—is then fitted to these perturbed samples to approximate the original model's decision boundary in the vicinity of the instance under analysis. The weights are assigned based on the proximity of the perturbed samples to the original instance, ensuring that the surrogate model prioritizes the local behavior of the black-box model. That is, the theoretical foundation of LIME treats the original model as a black box that maps inputs to outputs. By generating a neighborhood of perturbed samples and retrieving their corresponding predictions, LIME constructs a locally faithful model that reflects how the black-box model behaves near the instance of interest. This property, known as local fidelity, allows for meaningful explanations of individual predictions without requiring an understanding of the model's global structure or parameters. All data and models were hosted and estimated on the Center for Data Analysis and Supercomputing (CADS) at the University of Guadalajara through the use of the Leo Atrox Supercomputer.

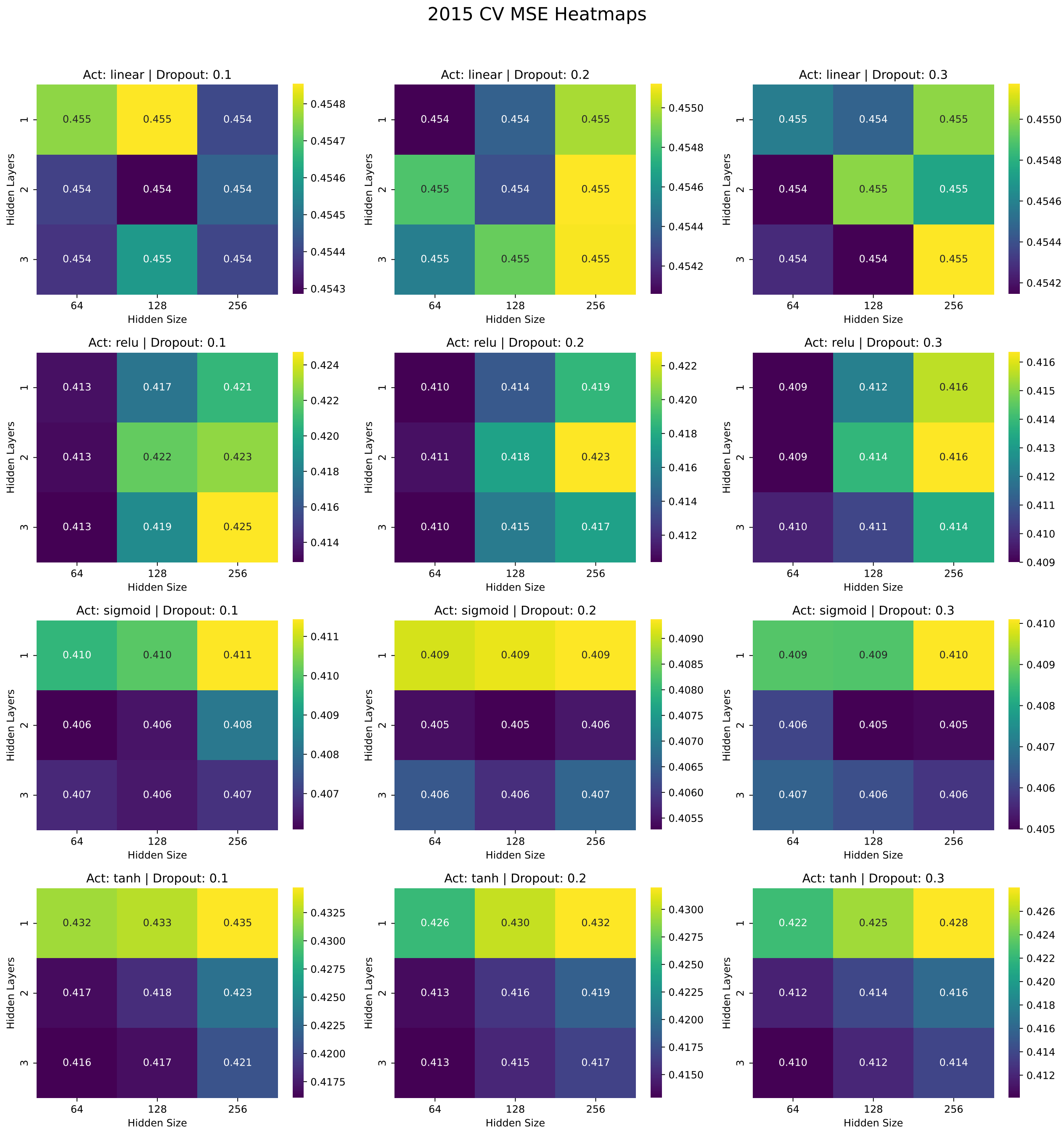
2.3 Output Analysis and Model Evaluation

2.3.1 2015 Cohort Network Specification

We trained several neural network models to predict elementary students' GPA based on survey responses. Using the cross-validated mean squared error (CV MSE), we evaluated models with varying configurations of hidden layer size, dropout rates, and number of layers. The Figure 1 presents the heatmaps for all models. On the top of each square its presented the dropout rates of 0.1, 0.2, and 0.3, respectively. The x-axis represents the hidden layer size with values of 64, 128, and 256 neurons, while the y-axis indicates the number of hidden layers ranging from 1 to 3. On the first line, we can find the linear activation function were the color scale range approximately from 0.454 to 0.455 and illustrates the MSE values, here darker colors indicate lower errors. The linear function with a dropout of 0.2, 2 hidden layers, and 128 neurons has a slightly better perfomance that other configurations, exhibiting the lowest MSE.

Following the heatmap, the second line corresponds to the relu activation function where the color gradient range from 0.414 to 0.424 and depicts MSE values with darker shades indicating lower errors, this models show that one or two hidden layers and small hidden layer sizes achieve a lower MSE compared to those with larger hidden layer. Following the heatmap, the third line corresponds to the sigmoid activation function, which MSE values spanning from 0.405 to 0.411 and darker shades indicating lower errors. In this specification, models with two or three hidden layers and moderate to larger hidden layer sizes generally achieve lower MSE values compared to those with a single hidden layer. We can see that configurations with dropout 0.3 and hidden layers sized 128 or 256 neurons show some of the best performances, with MSE values around 0.405. Overall, these results suggest that the sigmoid activation yields improved model performance compared to linear activation when employing deeper architectures with moderate dropout rates. Finally the fourth line corresponds to the tanh activation function,which MSE values span from 0.432 to 0.415 as before darker shades indicating lower errors.

Figure 1: 2015: Cross-Validation and Mean Square Error of Activation Functions by Dropout Rates.



In Table 1, we presents the mean and standart test scores and mean and standard train scores. The best-performing model uses an *sigmoid* activation function, a dropout rate of 0.3, and two hidden layers with 128 neurons each. This model achieved the highest cross-validated performance with a mean test score of -0.4049, indicating that it generalizes well and it does not presents overfitting. The low standard deviation of test scores (≈ 0.0065) across folds confirms that the model's performance is consistent across different data splits. The mean train score was -0.3726 and standart deviatiom traim score ≈ 0.005114). We can see that as a result of increasing the number of neurons per layer it helps to improved predictive accuracy when combined with dropout to prevent overfitting and data survey includes moderately strong predictors of GPA and that a well-regularized neural network can capture these patterns effectively.

Table 1: Best Models for each Hyperparameter Combination 2015.

Activation	Dropout	Hidden Size	Layers	Mean Test Score	Std Test Score	Rank	Mean Train Score	Std Train Score
sigmoid	0.3	128	2	-0.404970	0.006571	1	-0.372660	0.005114
sigmoid	0.3	256	2	-0.405063	0.006057	2	-0.381354	0.004388
sigmoid	0.2	128	2	-0.405264	0.007354	3	-0.375959	0.005906
sigmoid	0.2	64	2	-0.405418	0.006086	4	-0.375534	0.006936
sigmoid	0.2	256	2	-0.405516	0.006979	5	-0.378944	0.004998

The figure 2, it shows the evolution of training and validation losses (measured by Mean Squared Error, MSE) over successive epochs. The graph shows that both loss curves exhibit a rapid decline during initial epochs, this suggest that the model learns the structure of the data early in the training process and while the network is learning the training loss continues to decrease across all epochs. That is, the the network consistently improves its ability to fit the training data without encountering optimization instability or divergence. On the other hand, the validation loss decreases sharply at the beginning, reaching its minimum around epochs 12–15 and after this point, it plateaus and then increases slightly. The gap between the training and validation losses remains small and relatively constant throughout the learning process, pointed to a good generalization and the absence of overfitting, in addition to the smooth and monotonic behavior of the curves also suggests a stable optimization process and an appropriately chosen learning rate.

2.3.2 2015 Cohort Validation OLS

In addition to validate the results of these estimations, we performs a baseline OLS analysis to find if a linear regression model will estimate lower mean squared error that our the network estimation, we make this evaluation determinig a 5-fold cross-validation, the evaluation of better performance is as follows: If Baseline Test MSE $<$ Neural Network Test MSE, the OLS baseline predicts better than the neural network. If Baseline Test MSE $>$ Neural Network Test MSE, the neural network performs better than the linear baseline, capturing nonlinear effects, If Baseline Test MSE $\approx$ Neural Network Test MSE, the neural network and the linear baseline perform similarly, indicating that the problem is mostly linear.

Table 2 shows the comparison between the baseline OLS model and the neural network results. The OLS model achieves a cross-validation MSE of 0.4562 and a test MSE of 0.4479, indicating that it generalizes well without overfitting. The neural network achieves a lower train MSE of 0.3727 and a test MSE of 0.4050, with its cross-validation MSE closely matching

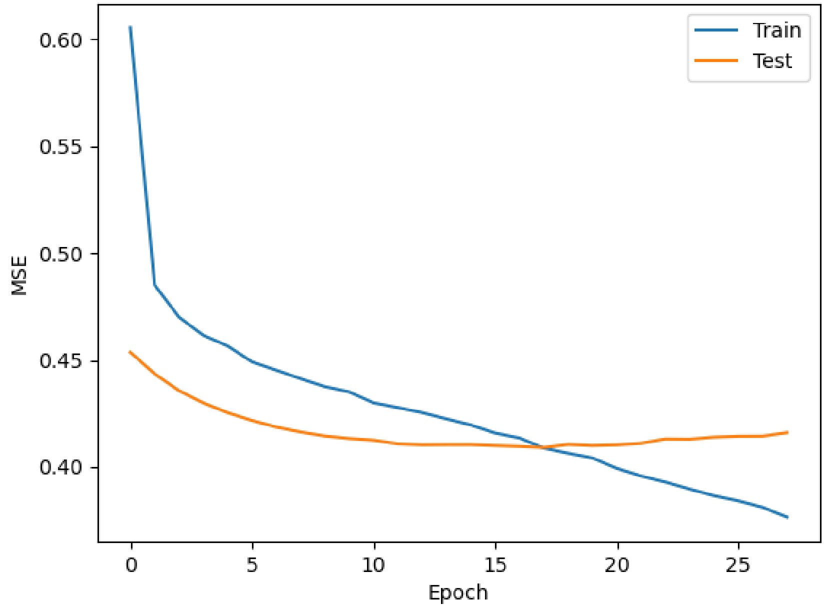


Figure 2: 2015: MSE loss and Epochs of Model with Sigmoid Act. Function [2 hidden layers and 128 neurons, 0.3 dropout rate]

Table 2: Comparison of Baseline OLS and Neural Network Performance (2015)

Model	**Train MSE**	**CV MSE**	**Test MSE**
Baseline OLS	–	0.4562	0.4479
Neural Network	0.3727	0.40497	0.4050

the test MSE (0.4050). This small difference between train, CV, and test errors suggests that the neural network is not overfitting and generalizes effectively. Since the test MSE of the neural network is lower than that of the OLS baseline, the neural network captures additional structure beyond the linear patterns, although the problem remains largely linear given the small improvement. Overall, the neural network slightly outperforms the linear baseline while maintaining good generalization.

2.3.3 2015 Model Interpretation Using LIME

The five best-performing models for each hyperparameter combination were retrained, after which the highest-ranked model (the black-box model) was selected and translated into an agnostic model (LIME) for further analysis. To facilitate the extraction of fetures from LIME, we construct from the survey structure a survey dictionary based on the survey components from the 2015 dataset, this allow us to link each feature to its corresponding survey construct. The Table 3 resumes the feature importance and survey section by showing the frequency of occuring (count) and mean importance. The indicator of mean importance is derived from the model's internal feature importance metric, that is each feature's relative contribution across the ensemble. The most importat features: `RFAB` (Family Resources & Well-being) exhibited the highest mean importance (0.1925), this is a strong positive contribution to model predictions. The feature sections `AB075` (Resources & Cultural Engagement) and `AB010` (Previous Education), showed importance values of 0.0861 and 0.0517, respectively.

Another features section such as `AB059` (Motivation & Self-Discipline) and `AB045` (Peer Interaction) display negative importance values, suggesting a modest diminishing effect on predicted outcomes. Some features that occurred frequently in the dataset, such as `AB067` and `AB034` had only moderate individual importance, whereas less frequent features could exert disproportionately strong effects, highlighting the nuanced interplay between feature prevalence and predictive influence.

In addition, the Table 4 reports a section-level summary obtained by aggregating the LIME importance values of individual variables. For each survey section, we computed the number of features, the total frequency with which those features appeared in the explanations, and descriptive statistics including the mean, (average importance of the variables within the survey section), maximum (highest importance value observed among the variables in the survey section.), and sum importance (sum of the importance values of all variables belonging to the survey section). These aggregated measures provide an overview of how features associated with each survey construct contributed to the model predictions. We can see that the feature section `Family Resources & well-being` exhibit the highest average importance and total sum of importance with only one feature (0.1925). The `Resources & Cultural Engagement` section(five features) displays a modest influece with a total importance of 0.1037 and average importance 0.0207. This result suggests that students' engagement with cultural or academic resources played a meaningful role in shaping the GPA as outcome. In contrast, sections such as `Peer Interaction`, `Identity`, `MARGINC` exhibited negative total importance (-0.0376, -0.0013, -0.0744). This implies that these features tend to lower the mode´s predicted scores. The `School Environment` and `Previous Educ` sections has the second and third feature usage count (98, 97) with a total importance of 0.0432 and 0.0339 suggesting that both supportive school conditions and previous education contribute moderate to predictions. Finally, the `Motivation & Self Discipline` section, represented by a single feature, shows a small negative importace of -0.00138. Because this section contains only a single feature, its average, maximum, and total importance values are identical. The result suggest that higher values of this feature slightly decrease predicted outcomes. However, the effect size is minimal and likely negligible in practice.

Table 3: 2015 Feature Importance and Survey Section Mapping

Feature	Count	Mean Importance	Survey Section
AB067	100	0.020877	20. Resources & Cultural Engagement
AB034	98	0.033983	17. School Environment
AB013	93	-0.008483	13. Previous Educ
AB060	28	0.058084	20. Resources & Cultural Engagement
RFAB	24	0.192507	30. Family Resources & well-being_index
AB059	17	-0.003130	14. Motivation & Self Discipline
AB075	15	0.086166	20. Resources & Cultural Engagement
AB002	6	-0.001386	11. Identity
AB010	4	0.051772	13. Previous Educ
AB079	2	-0.030038	20. Resources & Cultural Engagement
AB045	3	-0.0037669	18. Peer Interaction
AB074	1	-0.091445	20. Resources & Cultural Engagement
MARGINC	1	-0.074471	6. Demographics

Table 4: 2015: Aggregated Feature Importance by Survey Section

Survey Section	Total Features	Total Count	Avg Importance	Max Importance	Sum Importance
11. Identity	1	6	-0.001386	-0.001386	-0.001386
13. Previous Educ	2	97	0.021645	0.051772	0.043290
14. Motivation & Self Discipline	1	17	-0.003130	-0.003130	-0.003130
17. School Environment	1	98	0.033983	0.033983	0.033983
18. Peer Interaction	1	3	-0.037669	-0.037669	-0.037669
20. Resources & Cultural Engagement	5	147	0.020744	0.086166	0.103722
30. Family Resources & well-being_index	1	24	0.192507	0.192507	0.192507
6. Demographics	1	1	-0.001484	-0.074471	-0.074471

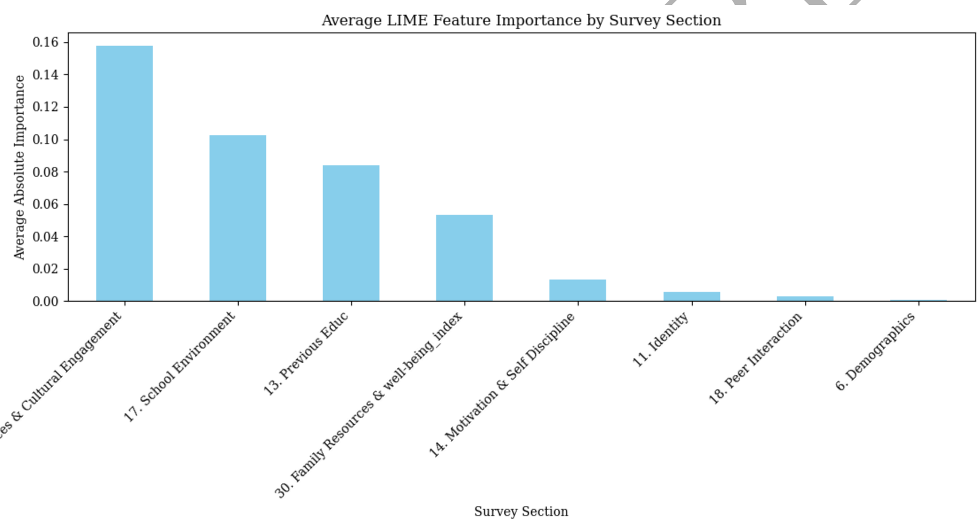


Figure 3: Average Lime Features by Survey Section 2015

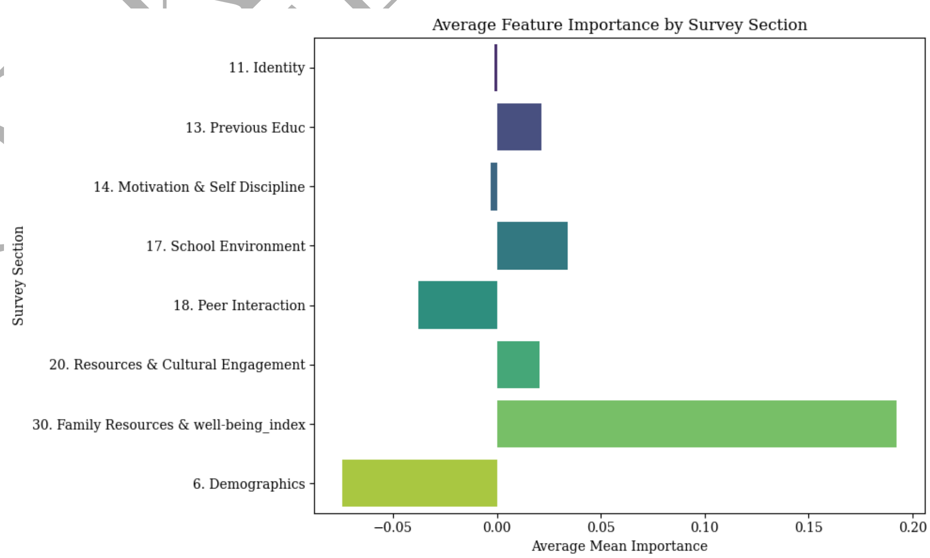


Figure 4: Average Feature Importance by Survey Section2015

2.3.4 2017 Cohort Network Specification

Similar to previous section, for the case of the 2017' cohort, we trained several neural network models to predict elementary students' GPA based on survey responses. By using cross-validation, we evaluated models with varying configurations of activation functions, hidden layer size, dropout rate, and number of layers. The Figure 5 presents cross-validated mean squared error (CV MSE) heatmaps for all different configurations in each row. The first row corresponds to the linear activation function, the MSE range from 0.444 to 0.443, darker colors indicate lower errors. For this activation function, the results show minimal variation in MSE across all configurations, suggesting that a stable model performance regardless of the dropout rate or network architecture within the tested parameter space.

The second row corresponds to the relu activation function where the color gradient span from 0.408 to 0.402.The heatmap for this activation function reveal that models with one or two hidden layers and small hidden layer sizes achieve lower MSE compared to those with larger hidden layer. The third row corresponds to the sigmoid activation function which MSE range from 0.402 to 0.399. The heatmaps reveal a consistent trend where models with two or three hidden layers and moderate to larger hidden layer sizes generally achieve lower MSE values compared to those with a single hidden layer. Finally, the fourth line corresponds to the tanh activation function which MSE range 0.4250 to 0.4075. Overall, the sigmoid activation funtion is consistently achieving the lowest CV MSE and it shows the most stable performance across architectures.

Table 5: Best Models for each Hyperparameter Combination 2017

Activation	Dropout	Hidden Size	Layers	Mean Test Score	Std Test Score	Rank	Mean Train Score	Std Train Score
sigmoid	0.3	128	2	-0.398629	0.002993	1	-0.373950	0.004144
sigmoid	0.3	256	2	-0.398890	0.003296	2	-0.376055	0.005556
sigmoid	0.2	128	2	-0.399143	0.002964	3	-0.375064	0.006486
sigmoid	0.2	128	3	-0.399257	0.003141	4	-0.367473	0.004633
sigmoid	0.3	64	2	-0.399516	0.003639	5	-0.369192	0.007110

In the table 5 we show the best-performing models for the 2017'cohort. The best-performing model, ranked 1st, uses a dropout rate of 0.3, a hidden layer size of 128, and 2 hidden layers. This model has the highest mean test score of -0.398629 and a stable mean train score of -0.373950, with a low standard deviation in training performance (0.004144). Compare to the second-ranked model, it uses a dropout rate of 0.3 and a higher hidden layer size of 256 with the same number of layers (2). Despite the bigger hidden size, this model performs worse on the test set, with a mean test score of -0.398890. The mean train score of -0.376055 is also slightly lower than that of the top model, although its performance on the training set is still stable. The third model uses a hidden layer size of 128 and a dropout rate of 0.2 with 2 hidden layers, and a mean test score of -0.395064. This performance is worse than the second model, though its mean train score of -0.375064 remains similar to that of the second model. Following the table, the fourth-ranked model uses a dropout rate of 0.2, which could imply fewer regularization during training. It also has a hidden layer size of 128 and 3 layers. While this model's performance on the test set (mean test score of -0.399257) is quite close to the other models, its mean train score of -0.367473 indicates that it is comparable to the others in terms of training performance, but the use of three layers doesn't significantly enhance the model's performance.

The fifth-ranked model, which uses a dropout rate of 0.3 and a hidden layer size of 64 with 2 layers, it has a mean test score of -0.399516. Its training performance (mean train score of -0.369192) is similar to the fourth-ranked model, but it is not enough to overcome the regularization effect of the dropout rate or achieve better generalization on the test set. In summary, the models with a dropout rate of 0.3 and varying hidden layer sizes and layer counts perform similarly with the best results coming from the model a hidden layer size of 128. While increasing the number of layers does not seem to drastically improve performance, dropout regularization plays a crucial role in helping the models generalize better, especially with the dropout rate of 0.3. However, it appears that models with a dropout rate of 0.2 did not perform as well, highlighting the potential importance of stronger regularization.

Table 6: Comparison of Baseline OLS and Neural Network Performance (2017)

Model	Train MSE	CV MSE	Test MSE
Baseline OLS	–	0.4421	0.4466
Neural Network	0.3739	0.3986	0.3986

Our next analysis is consider the behavior if the learning curve which is showed in Figure 6. In this graph we can see that the progression of both the training and validation losses over the course of 30 epochs. At the beginning of the training, the training loss (represented by the blue line) is relatively high but decreases during the first few epochs, indicating that the model is quickly learning and improving its fit to the training data. This drop is typical in the early stages of training, where the model adjusts its parameters and fit the data. After the initial drop, the training loss continues to decrease, but at a slower rate, indicating that the model is refining its learning and approaching an optimal solution (adam algorithm). Consequently, the validation loss decreases notably during the first several epochs and reaches its minimum around epochs 10–12. After this point, it stabilizes and displays small upward fluctuations, suggesting that the model's generalization performance stops improving beyond this stage. Comparing the two losses, the difference between training and validation losses remains small throughout most of the training process, suggesting good generalization and effective regularization, with only minor divergence toward the later epochs.The smooth behavior of both curves indicates that the optimization process is stable and that the chosen learning rate is appropriate In summary, since the validation loss ceases to improve meaningfully after approximately epoch 12 and starts to fluctuate slightly upward, early stopping around that point would likely provide the best balance between model performance and generalization.

2.3.5 2017 Cohort Validation OLS

Finally, the performance of the baseline OLS model and the neural network for 2017 is summarized in Table 6. The OLS model achieves a cross-validation MSE of 0.4421 and a test MSE of 0.4466, indicating that it generalizes well with no overfitting observed. The neural network achieves a lower train MSE of 0.3739 and a test MSE of 0.3986, with its cross-validation MSE closely matching the test MSE (0.3986). The small difference between training and test errors suggests that the neural network generalizes effectively without overfitting. Since the neural network achieves a lower test MSE than the linear baseline, it captures additional structure beyond the linear patterns.

Figure 5: 2017: Cross-Validation and Mean Square Error of Activation Functions by Dropout Rates.

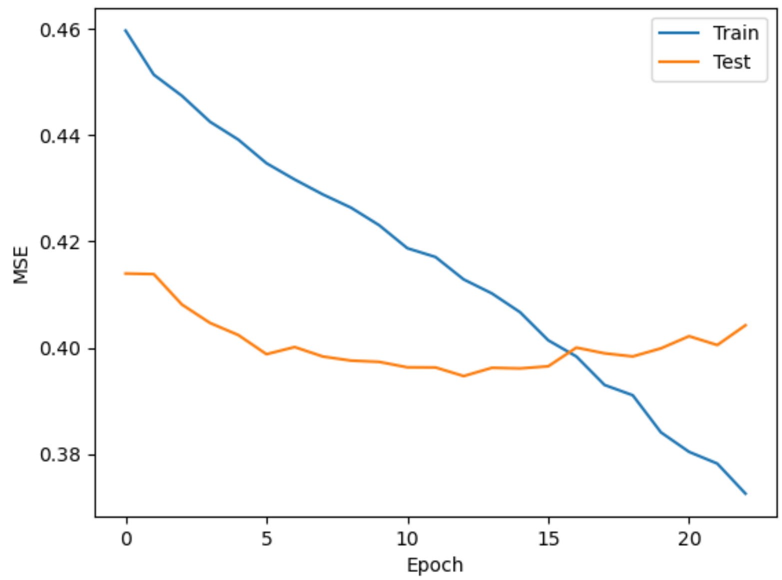


Figure 6: 2017: MSE loss and Epochs of Model with Sigmoid Act. Function [2 hidden layers and 128 neurons, 0.3 dropout rate]

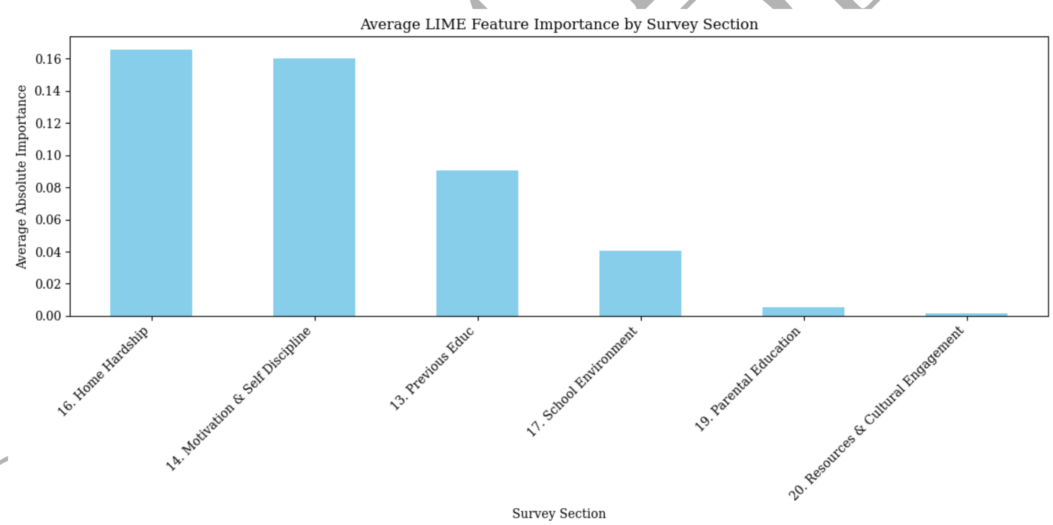


Figure 7: Average Lime Features by Survey Section 2017

2.3.6 2017 Model Interpretation Using LIME

The extraction of features via LIME algorithm follows the same procedude as previous section. The only difference is that the survey dictionary differs from 2015 in features, for this cohort was developed using questions from the 2017 dataset, linking each feature to its corresponding survey construct. This approach facilitated the interpretation of the LIME results within the context of relevant educational and socio-demographic factors, with the summarized findings presented in Tables 7 and 8 and Figures 7 and 8.

In the table 7 we present the features identified by the LIME. The results reveal that the most frequently occurring features are Motivation & Self-Discipline, Home Hardship, and Previous Education. For example, the features, PAB_54 (Motivation & Self-Discipline)

Table 7: 2017 Feature Importance and Survey Section Mapping

Feature	Count	Mean Importance	Survey Section
PAB_54	100	-0.008865	14. Motivation & Self Discipline
PAB_22	99	0.015030	16. Home Hardship
PAB_16	90	0.022243	13. Previous Educ
PAB_20	90	0.035708	16. Home Hardship
CTE	49	-0.001667	14. Motivation & Self Discipline
PAB_56	35	0.022300	17. School Environment
PAB_12	5	-0.105810	19. Parental Education
PAB_17	2	-0.004472	13. Previous Educ
PAB_74	2	0.091790	14. Motivation & Self Discipline
PAB_53	2	0.087217	20. Resources & Cultural Engagement
PAB_38	1	-0.089840	16. Home Hardship

Table 8: 2017: Aggregated Feature Importance by Survey Section

Survey Section	Total Features	Total Count	Avg Importance	Max Importance	Sum Importance
13. Previous Educ	2	92	0.008885	0.022243	0.017771
14. Motivation & Self Discipline	3	151	0.027086	0.091790	0.081258
16. Home Hardship	3	168	-0.013034	-0.035708	-0.039102
17. School Environment	1	35	0.022300	0.022300	0.022300
19. Parental Education	1	5	-0.105810	-0.105810	-0.105810
20. Resources & Cultural Engagement	1	2	0.087217	0.087217	0.087217

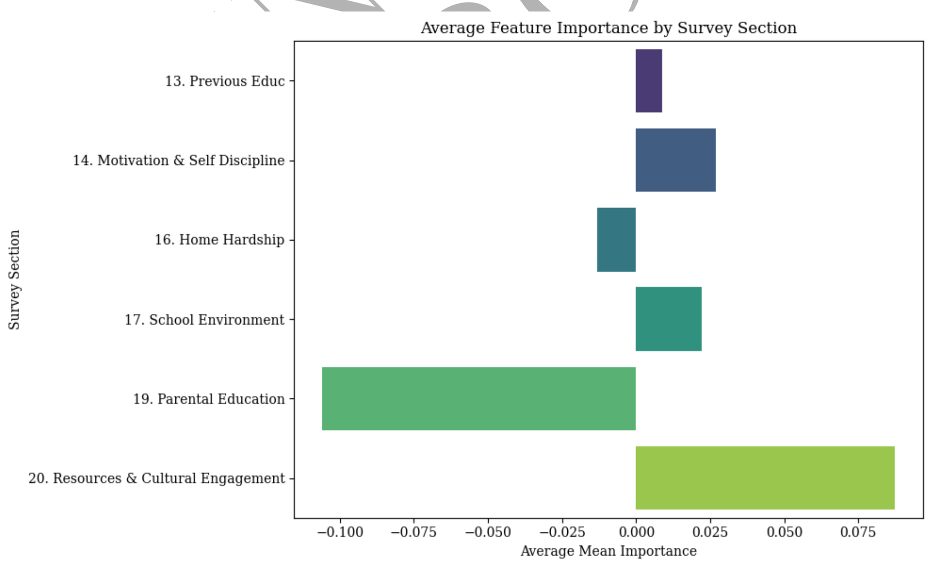


Figure 8: Average Feature Importance by Survey Section2017

appeared most frequently across 100 instances, though its mean importance score was slightly negative (-0.0088), suggesting that higher values of this variable tended to reduce the model's predicted outcome. PAB_22 (Home Hardship) showed a positive contribution (0.0150), and likewise, PAB_20 from the same section had a positive effect (0.0357), implying homogeneity in how aspects of home hardship influence the prediction. On the other hand, features related to Previous Education, particularly PAB_16 and PAB_17, showed some heterogeneity with importance values (0.0222 and -0.0044, respectively), indicating that educational background factors have different effect as predictors. With respect, parental education feature (PAB_12) it show negative contributions. Furthermore, some features exhibited extreme importance values—such as PAB_38 (-0.0808) and PAB_53 (0.0872)—suggesting that specific individual factors can have substantial localized effects on predictions, even if they are not consistently influential across all cases.

In the table 8 we presents the aggregated feature importance by survey section for the 2017 model. Resources & Cultural Engagement (Section 20) showed the highest individual feature importance (0.087217). Because this section contains only a single feature, its average, maximum, and total importance values are identical. This result suggests that students' engagement with cultural or academic resources played a meaningful role in shaping the model's predictions. This is supported by results of the section of motivation & Self-Discipline (three features) which displays strong influence, with a total importance of 0.081258 and a positive average importance (0.027086). These results highlight the consistent contribution of motivational and self-regulatory factors as predictors of student outcomes. In contrast, Home Hardship (Section 16), despite having the highest feature usage count (168), shows a negative total importance (–0.039102). This implies that higher levels of hardship tend to lower the model's predicted scores—which aligns with established evidence regarding the adverse effects of socio-economic disadvantage on educational achievement. The School Environment section has a small positive influence (0.022300), indicating that supportive school conditions contribute modestly to predictions. Finally the parental education section shows a strong negative importance (–0.105810), making it as an influential negative predictor. This suggests that lower parental educational attainment is associated with reduced predicted student outcomes, though the magnitude also indicates that this feature had a significant directional impact on the model. In general terms, the aggregated results show that cultural engagement and student motivation are the strongest positive contributors to the model's predictions, but, home hardship and parental education represent the strongest negative effects.

3 Discussion

The complex dimensions of factors that determine children's academic performance has been evaluated extensively in literature. The exploration of neural network settings will contribute to the area where from the mexican case analysis we can extract some highlights from our findings: First, from a behavioral perspective we observe that features related to home literacy practices (such as reading with family members (AB066, AB067), having books at home (AB062, PAB49) or parental engagement, (AB081-82, PAB80-84) are strong predictive importance and support the idea that enviromental passive "nudges" and household architectures shape children's learning behaviors [Thaler and Sunstein., 2008, Evans et al., 2010]. Second, in both years, features relate to school environment such as teacher-student interaction (AB034-56, PAB56-65), particularly student's perception of being heard by teachers, also emerge as influential, and peer-interaction are aligning with identity-based models where social affirmation enhances academic motivation [Akerlof and Kranton, 2002, Akerlof and Kranton, 2010, Cohen and Garcia, 2008].

Third, measures of socioeconomic marginality demonstrated negative associations with performance (MARGINC), also supported by the literature and with scarcity theory, which indicate that poverty imposes cognitive load and reduces attentional capacity

[Mani et al., 2013, Mullainathan and Shafi, 2013]. Furthermore, socioeconomic disadvantage (RFAB, PAB22, PAB38) consumes cognitive bandwidth, reducing the mental resources available for self-control and learning activities. Fourth, variables related to cultural participation (such as theater and fair attendance) were positive contributors, supporting theories of cultural capital and cognitive enrichment [Bourdie, 1986, Heckman, 2006]. Cultural activities provide counterbalancing cognitive stimulation, enriching students' mental models and supporting learning through increased curiosity and exposure. One of the features identified is previous education as a positive influece for academic performance , in [Shonkoff and Phillips, 2000, Heckman and Masterov, 2007], emphasizes that early childhood development is shaped by the interaction between biology and environment, including family background and public policy. They stresses the importance of high-quality, inclusive child care and thoughtful early interventions that support both cognitive and emotional development which are critical periods where small inputs yield long-term benefits.

4 Conclusions

The use of neural networks allows us to explore complex dimensions of factors that explain academic performance. In this paper, we propose "state of the art" methodological approach grounded in machine learning, using a feedforward neural network model. Our findings support the literature findings suggesting that children's academic achievement is the product of cumulative behavioral, environmental, and socioeconomic influences rather than cognitive ability alone. From a policy perspective, we considered that effective educational strategies should move beyond narrowly defined academic inputs. In both cohorts we see the need for policies that strengthen home learning environments. Features like reading at home and parental engagement in early literacy (Tables 3,4,7,8) can generate improvements in academic performance at relatively low cost. These interventions operate through passive behavioral channels, shaping learning habits without requiring substantial ongoing effort from households.

Another important feature is school environment represented by teacher support and school climate (AB034,AB045, PAB56). The promotion of inclusive classroom practices, teacher training focused on student voice, and supportive school environments may enhance students' academic motivation and persistence, particularly for those from disadvantaged backgrounds. Therefore, strengthening these relational aspects of schooling can complement traditional investments in curriculum and instructional quality. Furthermore, socioeconomic disadvantage (marginality) remains a significant constraint on academic performance, suggesting that educational policy cannot be fully effective in isolation. Interventions that reduce material hardship and psychosocial stress—such as school-based support services, nutritional programs, and income-related assistance—may indirectly improve learning outcomes by alleviating cognitive and emotional burdens associated with scarcity. This is actually a policy in course by government scholarships. The resources and cultural engagement features (AB067,AB074,AB060,AB075,AB079,PAB53) highlitht that by expanding access to cultural and enrichment activities will represent a promising alternative for policy intervention. The participation in cultural experiences through schools or community partnerships can provide cognitive stimulation and broaden learning opportunities, especially for students with limited access outside the school setting. Finally, the role of early childhood education (AB010,AB013,PAB16,PAB17) highlights the long-term returns to early intervention. Investments in high-quality, inclusive early childhood programs that support both cognitive and socio-emotional development are likely to yield sustained benefits across the life course.

Taken together, these findings in the Mexican context will require government actions that combine educational support with social protection programs given the persistent role of socioeconomic disadvantage observed in both cohorts.

References

[Abadi et al., 2015] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., and Zheng, X. (2015). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *Google Research*, pages 1–19.

[Ahmad and Shahzadi, 2018] Ahmad, Z. and Shahzadi, E. (2018). Prediction of students' academic performance using artificial neural network. *Computers and Education: Artificial Intelligence*, 40(3):157–164.

[Ahmetoglu et al., 2020] Ahmetoglu, E., Acar, I. H., and Asik Ozturk, M. (2020). Parental involvement and children's peer interactions. *International Journal of Educational Psychology*, 41:4447–4456.

[Akerlof and Kranton, 2002] Akerlof, G. A. and Kranton, R. E. (2002). Identity and schooling: Some lessons for the economics of education. *Journal of Economic Literature*, 40(4):1167–1201.

[Akerlof and Kranton, 2010] Akerlof, G. A. and Kranton, R. E. (2010). *Identity economics: How our identities shape our work, wages, and well-being.* Princeton University Press.

[Astone and McLanahan, 1991] Astone, N. M. and McLanahan, S. S. (1991). Family structure, parental practices, and high school completion. *American Sociological Review*, 56(3):309–320.

[Attewell, 2001] Attewell, P. (2001). The winner-take-all high school: Organizational adaptations to educational stratification. *Sociology of Education*, 74(October):267–295.

[Baashar et al., 2022] Baashar, Y., Alkawsi, G., Mustafa, A., Alkahtani, A. A., Alsariera, Y. A., Ali, A. Q., Hashim, W., and Tiong, S. K. (2022). Toward predicting student's academic performance using artificial neural networks (anns). *Applied Sciences*, 12(1289):1–16.

[Baker and Stevenson, 1986] Baker, D. P. and Stevenson, D. L. (1986). Mothers' strategies for children's school achievement: Managing the transition to high school. *Sociology of Education*, 59(3):156–166.

[Bandura, 1986] Bandura, A. (1986). *Social Foundations of Thought and Action: A Social Cognitive Theory.* Prentice-Hall.

[Baumrid, 1966] Baumrid, D. (1966). Effects of authoritative parental control on child behaviour. *Child Development*, 37:887–907.

[Bouchouna, 2024] Bouchouna, l. (2024). Enhancing children´s achievemet: The crucial role of school interaction and environmnet in learning. *Knowledge-International Journal,*, 63(2):205–209.

[Bourdie, 1986] Bourdie, P. (1986). *The forms of Capital.* Richardson, J., Handbook of Theory and Research for the Sociology of Education.

[Brooks-Gunn et al., 1997] Brooks-Gunn, J., Duncan, G. J., and Aber, J. L. (1997). *Neighborhood poverty: Context and consequences for children: Vol. 1.* New York: Russell Sage FoundationPress.

[Caldas and Bankston, 1997] Caldas, S. J. and Bankston, C. (1997). Effect of school population socioeconomic status on individual academic achievement. *The Journal of Educational Research*, 90(5):269–277.

[Caraway et al., 2003] Caraway, K., Tucker, C. M., Reinke, W. M., and Hall, C. (2003). Self-efficacy, goal orientation, and fear of failure as predictors of school engagement in high school students. *Psychology in the Schools*, 40(4):417–427.

[Cascallar et al., 2015] Cascallar, E., Musso, M. F., Kyndt, E., and Dochy, F. (2015). Modelling for understanding and for prediction/classification-the power of neural networks in research. *Frontline Learning Research*, 2(5):67–81.

[Cavanagh et al., 2006] Cavanagh, S. E., Schiller, K. S., and Riegle-Crumb, C. (2006). Marital transitions, parenting, and schooling: Exploring the link between family structure history and adolescents' academic outcomes. *Sociology of Education*, 79(4):329–354.

[Chollet, 2015] Chollet, F. (2015). Keras. URL https://github.com/fchollet/keras.

[Cohen and Garcia, 2008] Cohen, G. L. and Garcia, J. (2008). Identity, belonging, and achievement: A model, interventions, and implications. *Current Directions in Psychological Science*, 6(17):365–369.

[Coleman et al., 1966] Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., and York, R. L. (1966). *Equality of Educational Opportunity.* U.S. Government Printing Office.

[Davidson, 2019] Davidson, T. (2019). Black-box models and sociological explanations:predicting high school grade point average using neural networks. *Socius*, 5:1–11.

[Desforges, 2003] Desforges, C. (2003). *The Impact of Parental Involvement, Parental Support and Family Education on Pupil Achievement and Adjustment: A Literature Review.* Department for Education and Skills.

[Duckworth et al., 2007] Duckworth, A. L., Peterson, C., Matthews, M. D., and Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 92(6):1087–1101.

[Duckworth and Seligman, 2005] Duckworth, A. L. and Seligman, M. E. (2005). Self-discipline outdoes iq in predicting academic performance of adolescents. *Psychological Science*, 16(12):939–944.

[Duncan and Murnane, 2011] Duncan, G. J. and Murnane, R. J., editors (2011). *Whither Opportunity? Rising Inequality and the Uncertain Life Chances of Low-Income Children.* Russell Sage Foundation.

[Duncan et al., 1972] Duncan, O. D., Featherman, D. L., and Duncan, B. (1972). *Socioeconomic background and achievement.* NY: Seminar Press.

[Entwisle et al., 2005] Entwisle, D. R., Alexander, K. L., and Olson, L. S. (2005). *First Grade and Educational Attainment by Age 22: A New Story*, volume 110.

[Epstein and N., 1983] Epstein, J. and N., K. (1983). *The influence of friends on achievement and affective outcomes*. NY: Academic Press.

[Evans et al., 2010] Evans, M. D., Kelley, J., Sikora, J., and Treiman, D. J. (2010). Family scholarly culture and educational success: Books and schooling in 27 nations. *Frontiers in Education*, 28(2):171–197.

[Fite et al., 2014] Fite, P. J., Cooley, J. L., Williford, A., Frazer, A. L., and DiPierro, M. (2014). Parental school involvement as a moderator of the association between peer victimization and academic performance. *Children and Youth Services Review*, 44:25–32.

[Foster and Hagan, 2009] Foster, H. and Hagan, J. (2009). The mass incarceration of parents in america: Issues of race / ethninity, collateral damage to children and prisioner reentry. *The Annals of the American Academy of Political and Social Science*, 623:179–194.

[Gonida and Cortina, 2014] Gonida, E. and Cortina, K. S. (2014). Parental involvement in homework: Relations with parent and student achievement-related motivational beliefs and achievement. *British Journal of Educational Psychology*, 84:376–396.

[Hardgrave et al., 1994] Hardgrave, B., C, R. L. W., and Walstrom, K. A. (1994). Predicting graduate student success: A comparison of neural networks and traditional techniques. *Computers & Operations Research*, 21(3):249–236.

[Heckman, 2006] Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, 312(5782):1900–1902.

[Heckman and Masterov, 2007] Heckman, J. J. and Masterov, D. V. (2007). The productivity argument for investing in young children. *Review of Agricultural Economics*, 29(3):446–493.

[Hornik et al., 1989] Hornik, K., Maxwell, S., and Halbert, W. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366.

[Kahneman, 2011] Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

[Kingma P and Lei Ba, 2015] Kingma P, D. and Lei Ba, J. (2015). Adam:a method for stochastic optimization. *ICLR*, pages 1–15.

[Komarraju et al., 2011] Komarraju, M., Karau, S. J., Schmeck, R. R., and Alen, A. (2011). The big five personality traits, learning styles, and academic achievement. *Personality and Individual Differences*, 51:472–477.

[Komarraju et al., 2013] Komarraju, M., Ramsey, A., and Rinella, V. (2013). Cognitive and non-cognitive predictors of college readiness and performance: role of academic discipline. *Learning and Individual Differences*, 24:103–109.

[Lau et al., 2019] Lau, E. T., Sun, L., and Yang, Q. (2019). Modelling, prediction and classification of student academic performance using artificial neural networks. *SN Applied Sciences*, 1(9):982.

[LeCun et al., 2015] LeCun, Y. B., Yoshua, B., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

[Llorca et al., 2017] Llorca, A., Richaud, M. C., and Malonda, E. (2017). Parenting, peer relationships, academic self-efficay and academic achievement: Direct and mediating effects. *Frontiers in Psychology*, pages 1–11.

[Maehr, 2001] Maehr, M. L. (2001). Goal theory is not dead – not yet, anyway: A reflection on the special issue. *Educational Psychology Review*, 13:177–185.

[Mani et al., 2013] Mani, A., M., S., Shafir, E., and Zhao, J. (2013). Poverty impedes cognitive function. *Science*, 341(6149):976–980.

[Martin et al., 2013] Martin, A. J., Mansour, M., Anderson, M., Liem, G. A., and Sudmails, D. (2013). The role of arts participation in students' academic and nonacademic outcomes: A longitudinal study of school, home, and community factorst. *Journal of Educational Psychology*, 105(3):709–727.

[Molnar, 2022] Molnar, C. (2022). *Interpretable Machine Learning*. Leanpub, 2 edition.

[Mounts and Steinberg, 1995] Mounts, N. S. and Steinberg, L. (1995). An ecological analysis of peer influence on adolescent grade point average and drug use. *Developmental Psychology*, 31(6):915–922.

[Mueller and Parcel, 1981] Mueller, C. W. and Parcel, T. L. (1981). Measures of socioeconomic status: Alternatives and recommendations. *Society for Research in Child Development*, 52:13–30.

[Mullainathan and Shafi, 2013] Mullainathan, S. and Shafi, E. (2013). *Scarcity: Why having too little means so much*. Macmillan.

[Musso et al., 2020] Musso, M. F., Cascallar, E. C., Bostani, N., and Crawford, M. (2020). Identifying reliable predictors of educational outcomes through machine-learning predictive modeling. *Frontiers in Education*, 5(104):1–19.

[Musso et al., 2013] Musso, M. F., Kyndt, E., Cascallar, E. C., and Dochy, F. (2013). Predicting general academic performance and identifying the differential contribution of participating variables using artificial neural networks. *Frontline Learning Research*, 1(1):42–71.

[Nuñez et al., 2015] Nuñez, J. C., Suárez, N., Rosário, P., Vallejo, G., Valle, A., and L, E. (2015). Relationship between percieved parental involvement in homework, student homework behaviors, and academic achievement: differences among elementary, junior high, and high school students. *Metacognition Learning*, 10:375–406.

[OECD, 2013] OECD (2013). *PISA 2012 Results: Ready to Learn: Students' Engagement, Drive and Self-Beliefs (Volume III)*. OECD Publishing.

[Ribeiro et al., 2016] Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). "why should i trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1135–1144. ACM.

[Rodgers and Rose, 2001] Rodgers, K. B. and Rose, H. A. (2001). Personal, family and school factors related to adolescent academic performance. *Journal of Marriage and Family Review*, 33(4):47–61.

[Rodríguez-Hernández et al., 2021] Rodríguez-Hernández, C. F., Musso, M., Kyndt, E., and Cascallar, E. (2021). Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation. *Computers and Education: Artificial Intelligence*, 2(100018):1–14.

[Roksa and Potter, 2011] Roksa, J. and Potter, D. (2011). Parenting and academic achievement: Intergenerational transmission of educational advantage. *Sociology of Education*, 84(4):299–321.

[Rubinsteyn and Feldman, 2016] Rubinsteyn, A. and Feldman, S. (2016). fancyimpute: An imputation library for python. *URL https://github. com/iskandr/fancyimpute*.

[Rumbaut, 2005] Rumbaut, R. G. (2005). Turning points in the transition to adulthood: Determinants of educational attainment, incarceration, and early childbearing among children of immigrants. *Ethnic and Racial Studies*, 28(6):1041–1086.

[Shonkoff and Phillips, 2000] Shonkoff, J. P. and Phillips, D. A. (2000). *From Neurons to Neighborhoods: The Science of Early Childhood Development*. Washington (DC): National Academies Press (US).

[Sirin, 2005] Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review. *Review of Educational Research*, 75(3):417–453.

[Srivastava et al., 2014] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from over tting. *Journal of Machine Learning Research*, (15):1929–1958.

[Sun and Li, 2011] Sun, Y. and Li, Y. (2011). Effects of family structure type and stability on children's academic performance trajectories. *Journal of Marriage and Family*, 73:541–556.

[Thaler and Sunstein., 2008] Thaler, R. H. and Sunstein., C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness.* Yale University Press.

[Thomson et al., 1994] Thomson, E., Hanson, T. L., and McLanahan, S. S. (1994). Family structure and child well-being: Economic resources vs parental behaviors. *Social Forces*, 73(1):221–242.

[Tucker-Drob and Bates, 2015] Tucker-Drob, E. M. and Bates, T. C. (2015). Large cross-national differences in gene x socioeconomic status interaction on intelligence. *Psychological Science*, 27(2):138–149.

[Tucker-Drob and Harden, 2012] Tucker-Drob, E. M. and Harden, K. P. (2012). Learning motivation mediates gene-by-socioeconomic status interaction on mathematics achievement in early childhood. *Learning and Individual Differences*, 22:37–45.

[Vellymalay, 2012] Vellymalay, S. K. (2012). Parental involvement at home: Analyzing the influence of parents´socioeconomic status. *Studies in Sociology of Science*, 3(1):1–6.

[Wang and Degol, 2016] Wang, M.-T. and Degol, J. L. (2016). School climate: a review of the construct, mesurement, and impact on student outcomes. *Educational Psychology Review*, 28:315–352.

[Wang and Sheikh-Khalil, 2014] Wang, M.-T. and Sheikh-Khalil, S. (2014). Does parental involment matters for student achievement and mental health in high school. *Child Development*, 85:610–625.

[West et al., 2016] West, M. R., Kraft, M. A., Finn, A. S., Martin, R. E., Duckworth, A. L., Gabrieli, C. F., and D.E., G. J. (2016). Promise and paradox: Measuring students' non-cognitive skills and the impact of accountability. *Journal of Educational Psychology*, 38(1):148–170.

[Wolters, 1999] Wolters, C. A. (1999). The relation between high school students' motivational regulation and their use of learning strategies, effort, and classroom performance. *Learning and Individual Differences*, 11(3):281–299.

[Zimmerman, 2002] Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, 41(2):64–70.

Acknowledgments

The authors gratefully acknowledge the support provided to the project "Evaluation of the Relationship Between Childhood Obesity and Academic Performance: An Analysis Using Neural Networks" by the Center for Data Analysis and Supercomputing (CADS) at the University of Guadalajara through the use of the Leo Atrox Supercomputer.