AMPLIFYING THE PREDICTION OF TEAM PERFORMANCE THROUGH SWARM INTELLIGENCE AND MACHINE LEARNING

A Thesis

presented to

the Faculty of California Polytechnic State University,

San Luis Obispo

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Computer Science

by

Erick Harris

December 2018

© 2018 Erick Harris ALL RIGHTS RESERVED

COMMITTEE MEMBERSHIP

TITLE: Amplifying the Prediction of Team Performance through Swarm Intelligence and Machine Learning

AUTHOR: Erick Harris

DATE SUBMITTED: December 2018

COMMITTEE CHAIR: Franz Kurfess, Ph.D. Professor of Computer Science

COMMITTEE MEMBER: Hisham Assal, Ph.D. Professor of Computer Science

COMMITTEE MEMBER: David Askay, Ph.D. Professor of Communication Studies

ABSTRACT

Amplifying the Prediction of Team Performance through Swarm Intelligence and Machine Learning

Erick Harris

Modern companies are increasingly relying on groups of individuals to reach organizational goals and objectives, however many organizations struggle to cultivate optimal teams that can maximize performance. Fortunately, existing research has established that group personality composition (GPC), across five dimensions of personality, is a promising indicator of team effectiveness. Additionally, recent advances in technology have enabled groups of humans to form real-time, closed-loop systems that are modeled after natural swarms, like flocks of birds and colonies of bees. These Artificial Swarm Intelligences (ASI) have been shown to amplify performance in a wide range of tasks, from forecasting financial markets to prioritizing conflicting objectives. The present research examines the effects of group personality composition on team performance and investigates the impact of measuring GPC through ASI systems. 541 participants, across 111 groups, were administered a set of well-accepted and vetted psychometric assessments to capture the personality configurations and social sensitivities of teams. While group-level personality averages explained 10% of the variance in team performance, when group personality composition was measured through human swarms, it was able to explain 29% of the variance, representing a 19% amplification in predictive capacity. Finally, a series of machine learning models were applied and trained to predict group effectiveness. Multivariate Linear Regression and Logistic Regression achieved the highest performance exhibiting 0.19 mean squared error and 81.8% classification accuracy.

ACKNOWLEDGMENTS

Thanks to:

- My advisor, Franz Kurfess, for being my artificial intelligence mentor throughout my graduate program.
- David Askay and Lynn Metcalf for their continued guidance over the course of this project.
- Louis Rosenberg, Alex McClure, and Gregg Willcox at Unanimous A.I. for this incredible opportunity.
- My family for their constant support, advice, and encouragement.
- Nycole for her unconditional love, loyalty, and belief in me.

TABLE OF CONTENTS

			I	Page
LI	ST O	F TAB	LES	viii
LI	ST O	F FIGU	JRES	ix
Cl	НАРТ	TER		
1	Intro	oduction	n	1
	1.1	Motiva	ation	2
	1.2	Staten	nent of Hypothesis	3
	1.3	Contri	butions	3
2	Back	ground	1	4
	2.1	Swarm	1 Intelligence	4
		2.1.1	Swarms in Nature	4
			2.1.1.1 Colonies of Ants	5
			2.1.1.2 Flocks of Birds	5
			2.1.1.3 Swarms of Honey Bees	8
		2.1.2	Human Swarming	9
			2.1.2.1 Swarm AI	10
		2.1.3	Wisdom of the Crowd	14
	2.2	Group	Performance	16
		2.2.1	Group Personality Composition	17
		2.2.2	Social Sensitivity	19
	2.3	Machi	ne Learning	20
		2.3.1	Regression	22
		2.0.1	2.3.1.1 Multivariate Linear Regression	 ??

		2.3.1.2	K-Nearest Neighbor	. 23
		2.3.1.3	LASSO Regression	. 24
	2.3.2	Classific	cation	. 25
		2.3.2.1	Logistic Regression	. 25
		2.3.2.2	Decision Trees	. 25
		2.3.2.3	Naive Bayes	. 27
	2.3.3	Artificia	al Neural Networks	. 27
3	Experimen	tal Design	n	. 29
	3.1 Data	Collection	n	. 29
	3.2 Data	Analysis		. 31
	3.3 Mach	ine Learn	ing Pipeline	. 32
4	Results			. 34
	4.1 Perso	nality Fac	ctors	. 34
	4.2 Perso	nality Fac	cets	. 34
	4.3 Indiv	idual Min	imums and Maximums	. 35
	4.4 Socia	l Sensitivi	ity	. 38
	4.5 Mach	ine Learn	ing	. 38
5	Future Wo	rk		. 45
	5.1 Perso	nality Me	easurement	. 45
	5.2 Swarn	m Behavio	or	. 45
	5.3 Mach	ine Learn	ing	. 45
6	Conclusion	1		. 47
BI	BLIOGRAI	PHY		. 48

LIST OF TABLES

Table		Page
2.1	Performance Variance Explained by Neuman, Wagner, and Chris- tiansen (1999)	. 19
3.1	Gender Demographics	29
3.2	Race and Ethnicity Demographics	30
3.3	Age Demographics	30
3.4	Major Demographics	31
4.1	FFM Personality Factors Results	35
4.2	FFM Personality Facets Results	36
4.3	Individual Personality Factor Minimums and Maximums Results	37
4.4	Individual Personality Facet Minimums and Maximums Results	37
4.5	Reading the Mind in the Eyes Results	38
4.6	Features Chosen By Correlation Coefficients	39
4.7	Features Chosen By LASSO	40
4.8	Mean Squared Error for Regression Models	40
4.9	Mean Squared Error for Regression Models	41
4.10	Accuracy for Classification Models	42
4.11	Grid Search Cross Validation	43

LIST OF FIGURES

Figure		Page
2.1	Shortest Path for Ant Colony Optimization	. 6
2.2	Bird Flocking Social Rules	. 7
2.3	Wind Resistance in Flocks of Birds	. 8
2.4	Honey Bee Waggle Dance	. 9
2.5	Human Swarm using Swarm AI	. 10
2.6	Step by Step Swarm Convergence	. 12
2.7	Human Swarm predicting the Rose Bowl	. 13
2.8	Human Swarm predicting the EPL	. 13
2.9	Reading the Mind in the Eyes Sample Question	. 20
2.10	Linear Regression	. 23
2.11	K-Nearest Neighbor Regression	. 24
2.12	Logistic Regression	. 26
2.13	Decision Tree Example	. 26
2.14	Artificial Neural Network	. 28
3.1	Machine Learning Pipeline	. 32
4.1	Mean Squared Error for Regression Models	. 41
4.2	Mean Squared Error for Regression Models	. 42
4.3	Accuracy for Classification Models	. 43
4.4	Learning Curve For Artificial Neural Network	. 44

Chapter 1

INTRODUCTION

In 2016 an artificial intelligence program called AlphaGo defeated the best Go player in the world, Lee Sedol, 4 out of 5 games [63]. The program did not learn by mastering the rules of the game, but instead by studying 30 million moves made by expert players. It learned how humans make moves, what mistakes they make, and what opportunities they miss. The AlphaGo system was able to correctly predict Lee Sedol's next move 57% of the time [53]. This is concerning because we will always be at a disadvantage against artificial intelligences that are thoroughly trained to outperform us. Imagine if we could accurately predict what a person would do 57% of the time while developing consumer products, negotiating business deals, or pushing political agendas?

Moreover, these technologies will eventually be applied to industries that are more significant to our lives than games, and humans need to stay one step ahead in the intelligence "arms race". Various approaches to amplify our intelligence are being explored, ranging from genetic modification to brain-computer interfaces. However, other species in nature have faced similar challenges during their evolutionary development where their survival required a boost in their collective intelligence. Species like birds, ants, and honey bees have been able to amplify their intelligences by forming closed-loop systems that tap their combined knowledge and instincts. The principle behind these real-time systems is called Swarm Intelligence [55].

In recent years, technology has enabled groups of humans to form closed-loop systems that are modeled after natural swarms. Referred to as Artificial Swarm Intelligence (ASI), these methods have been shown to amplify the collective intelligence of groups, well beyond the capabilities of individual members [60]. This approach is sometimes called "blended intelligence" because it combines the computational efficiencies provided by AI systems with populations of real people [49]. Swarming keeps human intuition as an integral part of the processes. As a result, this "human-in-the-loop" approach combines the benefits of technology with the unique values that humans contribute like creativity, empathy, and morality.

1.1 Motivation

According to the Society for Industrial and Organizational Psychology, 13% of all employers, including 89 of the Fortune 100 companies, use personality tests to make hiring decisions and formulate business teams [16]. A modern organization's ability to quickly form effective groups to solve problems is critical to its success. The purpose of this study is to advance the understanding of team diversity as it relates to group performance and to investigate the potential advantages of measuring group personality through human swarms.

Furthermore, this study can better inform supervisors about the effects of personality composition on team performance. Two significant managerial benefits are evident. The first is the ability to survey the personality traits of potential group members before the actual creation of the team. With this information, the coordinator would be able to maximize the positive group characteristics identified in this study, and avoid the negative combinations. Additionally, supervisors can use this research to understand group personality compositions of existing project teams in order to manage them more effectively.

1.2 Statement of Hypothesis

Extant research on group performance suggests that group personality composition, measured through the aggregation of individual group members' personalities, will have a significant effect on team performance. Additionally, based on the predictive amplification demonstrated by human swarming, we expect that group personality composition, measured through swarm intelligence interfaces, will have an even stronger relationship with team performance.

1.3 Contributions

The research presented in this paper offers three primary contributions. First, the study advances the budding understanding of human swarming and demonstrates novel predictive capabilities of collective behavior. Second, this research heeds calls from previous studies to further examine the effects of group personality composition on team performance. Finally, this study explores the effectiveness of several widely-used machine learning techniques and compares their performance within a specific domain.

Chapter 2

BACKGROUND

2.1 Swarm Intelligence

Swarm Intelligence is an emerging field of biologically-inspired Artificial Intelligence based on the behavioral models of social animals like colonies of ants, flocks of birds, and schools of fish. These groups, or swarms, make a unified effort to amplify their collective intelligence by working together in real-time, closed-loop systems that converge on optimal solutions in synchrony. For example, honey bees use swarm intelligence to amplify their decision-making abilities while ants use swarm intelligence to improve their search and foraging efficiency.

2.1.1 Swarms in Nature

Artificial Intelligence research started in the 1950's by modeling intelligences found in nature. The first biologically inspired model was the Perceptron which led to the Artificial Neuron which lead to Neural Networks. It wasn't until 1989 when "Swarm Intelligence" was first coined by Gerardo Beni and Jing Wang. They defined it as "the collective behavior of decentralized, self-organized systems, natural or artificial" [1] [43]. Swarm Intelligence algorithms are modeled after social species found in nature. Evolved species like ants, birds, and bees have been able to amplify their collective intelligence by forming unique systems that tap their combined knowledge, wisdom, and instincts.

2.1.1.1 Colonies of Ants

One of the most well-known examples of swarm intelligences found in nature are colonies of ants. These nearly-blind insects work together to form a collective intelligence that is able to effectively determine the shortest path to nearby sources of food (F) as illustrated in Figure 2.1. Hundreds of ants disperse from the nest (N) and search for nourishment in separate directions. As the ants explore the surrounding area they leave a volatile chemical substance on the ground called pheromone. After the ants discover food sources they begin to travel back to the nest. The ants that discover the shortest path are able to make more trips between the source of food and the colony, leaving a denser trail of chemicals along the way. The other ants will begin to detect the heavy buildup of pheromone and they will redirect their foraging efforts accordingly [1].

The combined efforts of the colony have allowed ants to efficiently search for sources of food for tens of millions of years. This has allowed the species to survive and adapt to changing and often harsh environments in a manner that is incomparable to an individual ant. After studying the behavior of natural ant colonies Marco Darigo proposed the first Ant Colony Optimization (ACO) algorithm in 1991. This algorithm and its variants have been used to solve the Traveling Salesman Problem as well as other optimization, scheduling, and routing challenges [1] [43].

2.1.1.2 Flocks of Birds

Another social species that demonstrates collective intelligence in nature are flocks of birds. These animals work together to maneuver in a unified, cooperative manner in order to evade predators and migrate long distances. For example, high speed feedback-control enables thousands of starlings to make precision hairpin turns in



Figure 2.1: Shortest Path for Ant Colony Optimization
[14]

winds gusting 40 miles per hour [51]. The local interactions among birds, as illustrated in Figure 2.2, are based on the nearest neighbor principle, where birds follow certain flocking rules to adjust their motion (i.e., position and velocity) without any central coordination. In 1986, bird flocking behaviour was first simulated by Craig Reynolds where he proposed three simple flocking rules [47]:

- Separation flock members avoid collisions with nearby flockmates based on their relative position.
- Alignment flock members steer towards the average heading of nearby flockmates.
- **Cohesion** flock members attempt to stay close to nearby flockmates by flying in a direction that keeps them closer to the centroid.

Researchers have also found that pelicans fly in a V-shaped pattern to conserve energy during long and difficult journeys [43]. As illustrated in Figure 2.3, when birds lift their wings to fly, the wind that flows over their wing tips gives the birds in the back



Figure 2.2: Bird Flocking Social Rules

[3]

extra lift, reducing the amount of energy they need to fly. When the bird at the head of the V becomes tired, it falls to the rear of the formation where there is the least amount of wind resistance. This rotation through various positions in the V maximizes the use of each bird's energy, allowing the flock to fly for longer periods without taking breaks.

In 1995 James Kennedy and Russel Berhart developed the Particle Swarm Optimization (PSO) algorithm based on the natural behaviors of flocks of birds and schools of fish [1] [43]. Within the algorithm, artificial particles are constructed to "fly" inside the search space of optimization problems, mimicking the physical behaviors of flocking birds. PSO algorithms are able to avoid local minimums and accurately search for globally optimal solutions.



Figure 2.3: Wind Resistance in Flocks of Birds

[23]

2.1.1.3 Swarms of Honey Bees

The final and most relevant example of a natural swarm intelligence is formed by colonies of honey bees. These evolved creatures use their collective intellect to amplify their decision-making abilities, specifically in the selection of a new hive location every Spring. From hollow trees to abandoned sheds, the colony considers dozens of sites within a 30-mile area, each with their own tradeoffs. They must factor in size, ventilation, insulation, predators, nearest sources of pollen and water, and dozens of other variables. Although the individual bees lack the capacity to make a decision that is this complex, when hundreds of bees pool their knowledge and experience, they evoke a collective intelligence that not only makes a decision, but finds an optimal solution 80% of the time [55].

First, hundreds of scouter bees are sent out to locate potential hive sites. The bees assess the benefits and risks and return to the hive to make a collective decision. The most experienced bees communicate to each other through a series of body movements called "waggle dances". They are able to exchange information through signals that are embedded in the bodily vibrations. As illustrated in Figure 2.4, the signals can even tell another bee the exact location of a site 3 miles away. The angle of the waggle dance indicates the direction of the new hive location, and the speed and intensity with which they vibrate indicates their level of support. The honey bees will even physically butt heads if they disagree with one another.



Figure 2.4: Honey Bee Waggle Dance

By combining these signals, hundreds of bees will vibrate together and engage in a multi-directional tug of war. Factions will form in support of different options until finally, a certain threshold of support is met within the group, and the entire colony converges on the optimal solution for the swarm.

2.1.2 Human Swarming

Bees, fish, and birds developed a swarm intelligence over millions of years of evolution. Human beings have not yet been able to establish the subtle connections that other swarm animals use to form closed feedback loops. Schooling fish detect vibrations in the water, flocking birds feel motions propagating through the air, and honey bees use a complex body movement called a waggle dance. Human communication, while

^[9]

powerful, is too slow to achieve similar real-time, closed-loop systems necessary for swarming. We speak about 40-60 bits of information per second, but our brains are capable of processing information at 400 billion bits per second. Human beings can swarm only if we develop technologies to fill in the missing pieces that evolution has not yet provided. More specifically, swarming can occur among groups of online users by closing the loop around populations of networked individuals [51].



Figure 2.5: Human Swarm using Swarm AI

[58]

2.1.2.1 Swarm AI

Swarm AI is a real-time online environment for human swarming developed by Unanimous A.I and modeled after the behavior of honey bees [50]. It is the computational glue that allows people to work together in real-time swarms that can integrate noisy evidence, weigh competing alternatives, and converge on final decisions. Swarms of online users can answer questions by collaboratively moving a graphical puck to select from a set of possible answers as illustrated in Figure 2.5. The puck is generated by a central server and modeled as a physical system with a defined mass, damping, and friction [52]. Each participant in the swarm connects to the server and is provided a controllable magnet that allows users to apply force vectors on the puck simultaneously. Through the collaborative control of the puck, a real-time physical negotiation emerges among the members of the swarm. This occurs because all of the participants are able to push and pull on the puck at the same time, collectively exploring the decision-space, and converging upon the most agreeable answer, as depicted in Figure 2.6.

The puck moves in response to a dynamic feedback loop that is closed around all swarm participants. In this way, real-time synchronous control is enabled across all users, allowing them to negotiate and reach decisions in parallel. In the Swarm AI interface, as the magnet gets closer to the puck it applies a stronger force vector. If an individual wants to maintain a strong signal, it must closely follow the puck with its magnet. Participants who are confident in their answer will work harder to keep their magnets close, while others who aren't as confident may let their magnets drift away. All of this information is combined with Artificial Intelligence algorithms to drive the puck and make decisions [50].

Prior studies have shown that human swarms comprised of online participants have been able to produce significantly more accurate predictions than traditional methods like votes, surveys, and polls. For example, one study formed a human swarm of 75 amateur football fans and tested their ability to successfully predict the outcomes of college football games, as illustrated in Figure 2.7. The individuals by themselves averaged about 50% accuracy, but together as a swarm, they were able to achieve 70% accuracy against the Las Vegas spread, outperforming top football experts at ESPN [54]. Another study formed a human swarm using a group of volunteer soccer enthusiasts and analyzed their ability to predict English Premiere League soccer matches over a 5-week period. The results were compared to predictions made by the Sports Analytics Machine (SAM), a supercomputer developed by University of Sal-



Figure 2.6: Step by Step Swarm Convergence

[22]

ford that uses 10 years of data and extensive mathematical formulas to forecast EPL games. The individual participants averaged 55% accuracy and the supercomputer SAM averaged 64% accuracy [57]. However, the individuals were able to amplify their collective accuracy to 72% by forming real-time swarms, as shown in Figure 2.8.

In 2016 Unanimous A.I. was challenged to predict the Kentucky Derby. They gathered 50 horse-racing enthusiasts and had them form an artificial swarm online. Astonishingly the swarm correctly predicted the first four horses in order, a feat famously known as the Superfecta. Going off at 550 to 1 odds, anyone who placed a \$20 bet won \$11,000 [12]. Interestingly, not a single individual in the swarm got more than



Figure 2.7: Human Swarm predicting the Rose Bowl

[54]



Figure 2.8: Human Swarm predicting the EPL

[57]

1 of the 4 horses correct, but as a unified, coherent system, they were perfect. More recently, Unanimous AI used their Swarm Intelligence technology to predict the 2018

Oscars. A swarm of random movie-goers correctly predicted 15 of the 16 category winners. At 94% accuracy, this swarm comprised of average individuals out-predicted experts, movie critics, and a number of magazines [8]. While not a single participant had seen all of the movies, forming an artificial swarm intelligence allowed the group to pool their collective knowledge, experience, and intuition.

2.1.3 Wisdom of the Crowd

People are smart, but groups of people are even smarter. There has been over 100 years of research dating back to 1906 where Francis Gaulden famously studied the Wisdom of the Crowd (WOC) [50]. He attended a county fair in London and asked hundreds of farmers to estimate the weight of an ox. The individuals were wildly all over the place, but when he took the average of the responses it was remarkably accurate.

Typically, groups take a poll or a vote to capture the Wisdom of the Crowd. Votes and polls are valuable when it comes to characterizing populations, but they don't allow a group to express itself as a unified entity. Polls offer no means for populations to explore options and find consensus. Swarms, on the other hand, allow users to continuously update their intent in real-time, enabling groups to find commonality and converge on solutions that optimize support.

There's an old saying that says if one sheep jumps off a cliff, the entire herd will follow. This phenomenon has been observed in nature. If one member of the herd is startled and runs away, often the rest will run too. These types of herding systems amplify noise because one sheep can heavily influence the decisions made by the rest of the group. This instinctual behavior is designed to protect against predators, but it fails to breed intelligence. The decision-making process in herds is sequential - one member makes a decision and others follow suit. The scientific name for this group characteristic is social influence bias. *Social influence bias* is the idea that the aggregated opinions of others will have a significant influence on your own beliefs and actions [39]. Modern tools used for capturing the sentiments of a population, like online polling, amplify social influence bias and turn what could be a collective intelligence into a herd.

For example, product ratings on Amazon from other customers are visible to individuals leaving a new rating. If someone notices that hundreds of people have given a great review, it might influence how they rate the product. Facebook likes, Reddit up-votes, and Yelp reviews are all examples of online polls that exemplify social influence bias. A study at MIT created a voting system like Reddit where users can up-vote and down-vote in response to different posts. They found that the first vote had a 30% influence on how the next person would respond. By the end of the process, even after thousands of votes, the very first vote influenced the final decision by 25% [50]. While online polling is asynchronous and leads to herding, swarming is synchronous and nearly eliminates social influence bias. This is because every member of the swarm answers together, in real-time. Swarms capture the collective intelligence of the group, as opposed to capturing the intelligence of the first person to answer a poll.

Additionally, groups often fall victim to the *Tragedy of the Commons* - an economic theory where individuals who act according to their own self-interest behave contrary to the common good of all. For example, as an individual fisherman, it's entirely rational and moral to maximize the size of your catch. But, if all fishermen follow this prescription, the shared river gets depleted and is ruined for all. A demonstration was performed at the University of Maryland by social psychologist Dylan Selterman [56]. He posed an extra-credit challenge to his class, allowing each of his students to indicate privately how many points of extra credit they wanted on their exam - 2 points or 6 points. The only catch was that if more than 10% of the class asked for 6 points, then nobody would get any bonus. Clearly, it was in the best interest for everyone in the class to individually ask for 2 points, but that's not what happened. Instead, far too many students asked for 6 points and no one received extra credit.

A similar study was conducted by Unanimous A.I. in which 18 randomly selected users were asked to choose a cash bonus, the awarding of which was dependent upon the collective behavior of the group. The participants were asked to indicate on a blind survey whether they wanted a bonus of \$0.30 or \$0.90, but if more than 25% of the group asked for \$0.90, then none of the participants would get anything. As expected, 67% of the group asked for a \$0.90 bonus, well beyond the 25% threshold. However, the test was repeated by forming a real-time swarm, rather than taking a survey. The swarm configured itself such that 24% of the total pull on the puck was towards \$0.90, with 70% of the total pull towards \$0.30, and 6% abstaining [56]. The participants, while functioning as a unified swarm, avoided the Tragedy of the Commons, and instead converged on the solution that was best for the group as a whole. The Wisdom of the Crowd, while effective at fostering collective intelligence, appears to face some challenges that swarms naturally overcome.

2.2 Group Performance

Humans have been working in teams and organizing group tasks for thousands of years, as evidenced by the hundreds of miles of mounds stretching from Wisconsin to Florida built by Native American tribes around 2000 BC [10]. These elaborate manmade structures required moving millions of pounds of soil by teams using only small hand-held baskets. In the modern world, the use of teams in organizations has become so common that they are considered ubiquitous [61]. Companies increasingly rely on groups of individuals to reach organizational goals and objectives. However, many organizations struggle to cultivate optimal teams that can maximize performance [10].

Regardless of all the promise shown by utilizing groups, research has only seen moderate progress in understanding the reasons why some teams succeed and others fail. This challenge has led to broad research on group diversity traits including demographics, intelligence, and personality. One promising area of research in team performance is group personality composition (GPC) which concentrates on the configuration of personality traits in individual team members and the corresponding effect on group performance [10]. The current method of capturing GPC involves the adoption of the well-established Five-Factor Model of personality. Several studies have indicated a significant and considerable relationship between this set of personality traits and the resulting groups' performance [7] [21] [42] [6]. Prior research has also indicated that the average social sensitivity of the group is a strong predictor of team performance [70].

2.2.1 Group Personality Composition

Team diversity can be broken down into two dimensions: demography and process [64]. *Demography* includes the aggregated external characteristics of the group such as ethnicity, gender, and size. *Process* focuses on the teams' interactions and psychological dimensions, such as personality. The most popular personality model in both research and industry is the Five Factor Model (FFM). It originated in 1976 when Costa and McCrae used cluster analysis to identify three main categories of personality: extraversion, neurotisicm, and openness [11]. Conscientiousness and agreeableness were later added in 1997 by Pervin and John [26]. Over time, the Five Factor Model has become the most accepted personality taxonomy among researchers and

clinical psychologists [10]. The five traits are characterized in the following manner [65]:

- Extraversion Highly extraverted individuals are assertive and sociable, rather than quiet and reserved.
- Agreeableness Agreeable individuals are cooperative and polite, rather than antagonistic and rude.
- **Conscientiousness** Conscientious individuals are task-focused and orderly, rather than distractible and disorganized.
- **Neuroticism** Neurotic individuals are prone to experiencing negative emotions, such as anxiety, depression, and irritation, rather than being emotionally resilient.
- **Openness** Highly open individuals have a broad range of interests, are sensitive to art and beauty, and prefer novelty to routine.

Group personality composition (GPC) is measured by the group-level averages of each of the five personality factors as well as the individual-level minimum and maximum scores [10]. In 1998, a study was conducted that examined the correlation between all five GPC traits and team performance. Data was collected from 51 existing work teams, averaging roughly 13 employees per group. Team performance was captured by supervisor ratings over several dimensions of team effectiveness. The findings suggested that group-level means for conscientiousness, agreeableness, and neuroticism all had a significant relationship with team performance [6]. In another study, 82 teams of retail associates were observed in different product category departments within a large retail chain. Group performance was measured by the number of customer complaints per month and evaluations from managers and supervisors. The authors found that all five personality factors were effective predictors of team performance. Together, the GPC traits explained 29% of the difference in team effectiveness between groups [42]. The breakdown of the variances explained by each category is listed in Table 2.1 [41]. In this study, group personality composition will be calculated through the group-level averages of the Big Five Inventory (BFI) personality test, which is a widely-accepted measure of the Five Factor Model [27] [25] [24]. Agreeableness, Conscientiousness, and Extraversion are the personality dimensions that are expected to have the strongest relationship with group performance.

Factor	R^2
Extraversion	0.09
Conscientiousness	0.17
Agreeableness	0.18
Neuroticism	0.08
Openness	0.10

Table 2.1: Performance Variance Explained by Neuman, Wagner, and Christiansen (1999)

2.2.2 Social Sensitivity

Reading the Mind in the Eyes (RME) is a psychological assessment that measures the emotional intelligence, or "social sensitivity", of the respondent. Social sensitivity is often described as the ability to perceive, interpret, and respond to the intentions, dispositions, and behaviors of others [59]. These skills are extremely important for effective decision-making and problem-solving in groups, as understanding the needs, goals, and beliefs of others is a fundamental requirement of any collaborative team. The RME assessment includes 36 questions, each showing a cropped facial image restricted to a region around the eyes, and a set of four options that describe the emotion expressed. Participants are tasked with reading the emotional state of the facial image based on visual cues provided by the eyes [5]. Prior studies have shown that the RME test is a reliable measure of social sensitivity, with strong internal consistency and test-retest stability [59]. For example, in a study of 699 subjects, working in teams of two to five, average social sensitivity accounted for roughly 7% of the group's collective intelligence [70]. Based on these findings, the present study additionally measures the social sensitivities of groups as well as the personality compositions and analyzes the corresponding effects on team performance.



Figure 2.9: Reading the Mind in the Eyes Sample Question
[31]

2.3 Machine Learning

Artificial Intelligence (AI) is a broad discipline that encompasses several fields of study including machine learning, natural language processing, and computer vision. Machine Learning (ML) is a subset of AI that uses complex learning algorithms to discover patterns in large data sets, usually in order to make predictions or classifications about new observations. One of the earliest definitions of machine learning came from Arthur Samuel in 1959, who described it as a "field of study that gives computers the ability to learn without being explicitly programmed" [69]. Samuel created a program that played thousands of games of checkers against itself. The program started to recognize patterns that led to wins and patterns that led to losses. Machine learning algorithms, like Samuel's, construct a framework for learning. The training process commonly includes some method of trial and error with the ability to measure that error and make changes to improve the framework. Machine learning is divided into three distinct categories: supervised learning, unsupervised learning, and reinforcement learning.

Supervised learning is task driven and in supervised learning problems, the training data is labeled [37]. This means that every set of input features has an associated output value. As the model trains, it can measure the error in its predictions and make adjustments to the model. A well-known supervised learning problem involves predicting the value of a house based on features like age, size in square feet, and number of bedrooms. On the other hand, *unsupervised learning* is data driven, and the training data is not labeled [37]. Input features have no corresponding output values. Unsupervised learning algorithms are used to discover hidden structure or complex patterns within data sets. A well-known unsupervised learning task is called customer segmentation, where customers are clustered together based on high-dimensional similarities.

Finally, *reinforcement learning* algorithms are designed for agents that take actions within an environment. Over time these agents can learn to maximize reward signals and avoid negative feedback, eventually developing an optimal policy, or strategy, for interacting in that environment. Reinforcement learning algorithms are well suited for learning behavioral patterns. This study, however, is primarily concerned with supervised learning.

2.3.1 Regression

In supervised learning, the algorithm's goal is not to memorize the data, but instead to generalize it and develop a model that accurately reflects the underlying relationships within the observations. The first type of supervised learning, called Regression, uses input features to predict continuous independent variables, like the price of a house. Three common regression algorithms include Multivariate Linear Regression, K-Nearest Neighbor Regression, and LASSO Regression.

2.3.1.1 Multivariate Linear Regression

Multivariate linear regression (MLR) is one of the most widely used statistical techniques in academia and industry. It was defined in 1982 as a technique for determining the correlation between a response variable Y, and some combination of two or more predictor variables X_i [38]. MLR provides an estimate of both magnitude and statistical significance of relationships between variables. A multivariate linear regression model is defined by the equation: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k$ where Y is the response variable, X represents the predictor variables, and β delineates the population regression coefficients [15]. The regression equation, illustrated by the red line in Figure 2.10, can be used to make predictions which can be compared with the true value of the observation to produce an error. The model then adjusts the population regression coefficients β in a way that minimizes future error. With enough examples, the learning agent can eventually develop a formula that effectively models the true relationship between predictors and response variables.

One of the primary challenges with multivariate linear regression and other supervised learning techniques is determining which variables to include in the model and which to leave out. Clearly, all potential predictor variables could be included, but that often



Figure 2.10: Linear Regression

[18]

results in a cumbersome model that is difficult to interpret. On the other hand, a model that only includes a handful of predictor variables often provides substantially more accurate predictions [34]. The correlation coefficient, R, describes the strength of the relationship between two variables, and is often used to determine the best input features for the model [34].

2.3.1.2 K-Nearest Neighbor

K-Nearest Neighbor (KNN) Regression is a relatively simple supervised learning algorithm that utilizes feature similarity to develop its models. New observations are evaluated on their resemblance to previous observations, as illustrated in Figure 2.11. The distances between the new observation, which is denoted by the blue star, and the K-closest data points are measured, and averaged to produce a new prediction [33].

The performance of KNN regression is strongly dependent on the parameter K, as locality is defined in terms of the K samples nearest the input vector. Finding the best values of K is essential, as large K values can decrease the effect of noise and minimize prediction loss. However, small K values allow for simple implementation



Figure 2.11: K-Nearest Neighbor Regression

[28]

and efficient queries [17]. Grid search cross validation is an optimization technique that can be used to determine the best K value for the training set. KNN is one of the most influential regression algorithms with over 18,000 publications and has been successfully applied to time series forecasting in finance, climatology, and other domains [2].

2.3.1.3 LASSO Regression

LASSO is another regression method that performs both feature selection and regularization simultaneously. It was first formulated by Robert Tibshirani in 1996 and stands for Least Absolute Shrinkage Selector Operator [67]. This is because the algorithm applies a penalty to some of the input features and "shrinks" their coefficients to zero, effectively removing them from the model (regularization). Variables that still have a non-zero coefficient after the shrinking process are selected to be part of the model (feature selection) [36]. This helps reduce variance without substantial increase in bias which is especially useful in problems with a small number of observations but a large number of features. LASSO also eliminates irrelevant features that are not associated with the response variable which helps improve performance and reduce over-fitting [19].

2.3.2 Classification

The second type of supervised learning, called Classification, uses input features to classify categorical variables, like a "benign" or "malicious" tumor. Three common classification algorithms include logistic regression, decision trees, and naive bayes classifiers.

2.3.2.1 Logistic Regression

Logistic Regression is used for the analysis and prediction of dichotomous outcomes. It was proposed as an alternative to Ordinary Least Squares regression in the late 1960s and early 1970s [45]. Logistic regression is well suited for modeling relationships between a binary outcome variable and one or more predictor variables. As illustrated in Figure 2.12, the logistic regression algorithm determines the equation of the best fitting line to separate the two classes. When new observations are presented, the algorithm classifies them based on which side of the decision boundary they fall.

2.3.2.2 Decision Trees

The Decision Tree algorithm is a sequential approach to binary classification that utilizes a directed graph data structure. It was first proposed by Gordon V. Kass in 1980 and exhibits high classification speed, strong learning ability, and a relatively simple implementation [44]. It does this by determining which questions, or "tests",



Figure 2.12: Logistic Regression

[29]

best divide the observations into separate classes [32]. Figure 2.13 illustrates a rudimentary classification tree for predicting whether or not a golf tournament will be held, based on three predictor variables related to the weather: outlook, humidity, and wind. Each node of the tree describes a test that is conducted on the observation, and the leaves of the tree represent outcome predictions based on the results of those tests. Classification trees, while effectively comprehensible, are overly sensitive to features with a large number of values [40].



Figure 2.13: Decision Tree Example

[35]

2.3.2.3 Naive Bayes

Lastly, Naive Bayes is a probabilistic classification technique that is based on the Bayes Theorem and assumes that all input features are independent of one another [30]. Bayes Theorem describes the likelihood of an event through the prior probabilities of similar events occurring. Naive Bayes is particularly well suited for highdimensional data sets and has shown to outperform complex classification techniques, despite its simplicity [48]. It has proven effective in many practical applications, including text classification, medical diagnosis, and systems management.

2.3.3 Artificial Neural Networks

Artificial Neural Networks (ANNs) are more sophisticated computational models based on the structure and functionality of the human brain, which is composed of roughly 100 billion neurons [62]. Each neuron is connected to thousands of others through their dendrites (input nodes) and axons (output nodes). Similarly, an artificial neural network is made up of several layers of digital neurons, each of which is given input values from nodes in the preceding layer and generates an output to nodes in the following layer, as illustrated in Figure 2.14.

Each input node represents an independent, or predictor variable of the problem. For example, if the network were to estimate the value of a used car, then the input variables might include the vehicle's make, model, and mileage. The output node represents the dependent, or response variable(s) of the problem and can be continuous, binary, or categorical. Predictions are made through a process called Forward Propagation where activated neurons receive weighted inputs from other neurons in the preceding layer [62]. The weights are summed and fed into an activation function. The output of this function is then sent to nodes in the following layer. As the neural



Figure 2.14: Artificial Neural Network

[20]

network is trained, the weighted values between neurons are continuously adjusted until an optimal model, or mapping from inputs to outputs, is achieved [37]. Artificial neural networks have resurrected interest in Artificial Intelligence and have been used to create revolutionary technologies like self-driving cars, facial-identification software, and intelligent personal assistants like Apple's Siri and Amazon's Alexa.

Chapter 3

EXPERIMENTAL DESIGN

The scope of this study is focused on student teams working together towards common goals in time-constrained environments, where all team members face real consequences for either success or failure. These student groups participated in various business and engineering courses that focused on creating innovative products and services. Although there has been some criticism of student sample usage, it is frequently the sample of choice used in group personality composition research [10].

3.1 Data Collection

Data for this study was collected from 541 college students, across 111 teams, at a large American public university. Participants completed a set of well-accepted and vetted assessments that allowed for the quantification of each group's personality composition. After 12 weeks, course instructors completed performance evaluations for each team over several dimensions of group effectiveness. The overall survey response rate was 74% with individual demographic distributions listed in Tables 3.1, 3.2, 3.3, and 3.4.

Gender	Percent
Male	56.2%
Female	43.8%

 Table 3.1: Gender Demographics

Race and Ethnicity	Percent
American Indian or Alaska Native	0.2%
Asian	10.9%
Black or African American	0.7%
Hispanic/Latino American	9.9%
White	72.8%
Other	5.2%

Table 3.2: Race and Ethnicity Demographics

The independent variables in this study include the group's demographic configuration, personality composition, and levels of social sensitivity. Group demographic information, including team size, gender, and ethnicity were captured using popular survey instruments like Qualtrics. Personality characteristics were derived using a widely-accepted measurement of the Five Factor Model, called the Big Five Inventory (BFI) [27]. Each factor of the Big Five assessment breaks down into two supplementary facets, generating 10 additional personality categories to be evaluated [66]. Finally, the group's social sensitivity was captured using the Reading the Mind in the Eyes (RME) questionnaire.

Age	Percent
18-19	11.3%
20-21	57.0%
22-23	26.9%
24-25	4.8%

 Table 3.3: Age Demographics

Subjects were first asked to take the assessments individually to determine the grouplevel mean, minimum, and maximum scores for each of the personality factors. Then

Major	Percent
Business Administration	35.7%
Communication Studies	12.9%
Engineering	31.3%
Other	20.1%
Communication Studies Engineering Other	12.9% 31.3% 20.1%

 Table 3.4: Major Demographics

groups were instructed to repeat the assessments together as unified swarms using the Swarm AI technology provided by Unanimous A.I. These assessments included a 44-item personality test (BFI) and a 36-item social sensitivity test (RME).

The dependent variables in this study include several assessments of the team's performance by the professor in each course. Extant research has suggested the use of a second dependent variable to measure the same outcome phenomenon as the primary dependent variable [10]. Therefore, two assessments of group performance were administered. The first was a 6-item evaluation of team effectiveness and the second was a 4-item measure of team performance. Prior studies have shown that both assessments provide an accurate and reliable measure of group performance [68]. These scores were combined to create a single dependent variable that represents the overall performance of the group.

3.2 Data Analysis

A variety of statistical techniques were applied to analyze the effects of group composition on team performance. Most notably, the coefficient of determination, R^2 , was used to calculate the variance in group performance explained by a particular independent variable. T-Tests were conducted to measure the significance of these relationships. Additionally, the variances explained by group-level personality averages were compared to the variances explained by swarm-level personality scores. Furthermore, a mixture of regression and classification algorithms were deployed, including Multivariate Linear Regression, K-Nearest Neighbor Regression, LASSO Regression, Logistic Regression, Decision Trees, and Naive Bayes Classifiers. Finally, an artificial neural network was developed in TensorFlow and its performance was compared with the simpler machine learning models.

3.3 Machine Learning Pipeline



Figure 3.1: Machine Learning Pipeline

As illustrated in Figure 3.1, a machine learning pipeline was constructed to perform the complex computational analysis required for this study. First, raw survey responses from participants and instructors were exported from Qualtrics and Swarm AI. The spreadsheets were parsed and the survey responses were extracted and scored. Invalid teams were discarded, including groups that exhibited less than 50% survey completion or less than 50% swarm participation [10]. The scores were scaled to account for acquiescence, the tendency of an individual to consistently agree to questionnaire items, regardless of the content [46]. Then group-level personality compositions were calculated by taking the mean of the individual team members' scores. The correlation of determination, R^2 , was used to measure the variance in team performance explained by group-level and swarm-level personality factors.

Then the data was prepared for a series of machine learning algorithms. The correlation coefficient, *R*, was used to perform feature evaluation and determine the top 10 predictors of team performance. Next, the data was normalized, partitioned into training and testing sets, and fed into three regression algorithms. The top 10 feature model was compared with a more complex model that utilized the full set of predictor variables. Next, groups that performed in the top 75th percentile were labeled as "exceptional teams". Three separate classifying algorithms were trained to predict whether groups would fall under this category. Finally, an artificial neural network was constructed, trained, and validated over the data set. Network hyperparameters, including number of hidden layers, activation functions, batch size, and more were optimized using grid search. The various machine learning models were compared using Mean Squared Error (MSE) and classification accuracy (ACC) over the validation set.

Chapter 4

RESULTS

4.1 Personality Factors

This study examines the aggregated personality compositions of the group and investigates the variance in team performance explained by group-level means and swarm-level responses. As illustrated in Table 4.1, personality compositions measured by group averages, explained 8% of the variance in team performance. On the other hand, group personality compositions measured by swarms explained 15% of the variance. Thus, when GPC was determined as a swarm, rather than aggregating the individual member means, it increased predictive capabilities of team performance by 7%.

While prior studies have found Conscientiousness and Agreeableness to be the personality factors that best predict group performance, the results of this research suggest that Extraversion and Conscientiousness are the leading predictors. Additionally, GPC differences measured by the swarm were strongest in Extraversion (7%) and Openness (4%). This has interesting implications in the context of local interactions among swarm members and suggests that swarming more accurately captures social, active, and creative group personality compositions.

4.2 Personality Facets

The predictive capabilities displayed by human swarms is even more apparent when analyzing the 10 personality facets of the Five Factor Model, shown in Table 4.2.

Personality Factor	Group β	Swarm β	Group R	Swarm R	ΔR	Group \mathbb{R}^2	Swarm \mathbb{R}^2	ΔR^2
Extraversion	0.54	0.49	0.25	0.36	0.11	0.06*	0.13*	0.07
Agreeableness	0.55	0.23	0.14	0.12	-0.02	0.02	0.01	-0.01
Conscientiousness	0.71	0.38	0.20	0.24	0.04	0.04	0.06^{*}	0.02
Neuroticism	-0.19	-0.21	-0.05	-0.10	-0.05	0.00	0.01	0.01
Openness	0.21	0.61	0.05	0.21	0.16	0.00	0.04	0.04
All Factors			0.29	0.39	0.10	0.08	0.15*	0.07

*P-Value < 0.05

Table 4.1: FFM Personality Factors Results

While the aggregated individual means explained 10% of the variance in team performance, swarm-level personality scores explained 29% of the variance, representing a 19% increase in predictive capacity. Interestingly, it was common for one of the two personality facets to exhibit a stronger relationship with group performance than the other. For example, within the dimension of Agreeableness, Altruism accounted for 7% of the variance while Compliance exhibited no correlation with performance. These findings suggest that, of the personality facets measured in this study, Assertiveness (E), Activity (E), Altruism (A), Order (C), and Self-Discipline (C) are the traits that have the greatest effect on group performance. Additionally, swarm amplification across all personality dimensions was strongest in Extraversion (10%), Agreeableness (5%), and Neuroticism (5%).

4.3 Individual Minimums and Maximums

The individual minimum and maximum scores for each personality category were also recorded and analyzed. The results, shown in Tables 4.3 and 4.4, suggest that indi-

Personality Facet	Group β	Swarm β	Group R	Swarm R	ΔR	Group \mathbb{R}^2	Swarm \mathbb{R}^2	ΔR^2
Assertiveness (E)	0.48	0.48	0.22	0.36	0.14	0.05	0.13*	0.08
Activity (E)	0.48	0.31	0.24	0.28	0.04	0.06*	0.08^{*}	0.02
Altruism (A)	0.60	0.49	0.15	0.27	0.12	0.02	0.07^{*}	0.05
Compliance (A)	0.20	-0.05	0.06	-0.03	-0.09	0.00	0.00	0.00
Order (C)	0.50	0.30	0.21	0.29	0.08	0.04	0.08*	0.04
Self-Discipline (C)	0.69	0.29	0.21	0.18	-0.03	0.04	0.03	-0.01
Anxiety (N)	-0.03	-0.26	-0.01	-0.14	-0.13	0.00	0.02	0.02
Depression (N)	-0.55	0.07	-0.18	0.05	0.22	0.03	0.00	0.03
Aesthetics (O)	0.33	0.27	0.14	0.14	0.00	0.02	0.02	0.00
Ideas (O)	-0.13	0.33	-0.03	0.16	0.19	0.00	0.02	0.02
All Facets			0.32	0.54	0.22	0.10	0.29*	0.19

*P-Value < 0.05

Table 4.2: FFM Personality Facets Results

vidual outliers may play an even greater role in team performance than the average personality composition of the group. For example, after examining the impact of team conscientiousness, having a single high scoring member is more predictive of group performance (7%) than having a team with a high average level of conscientiousness (4%). This suggests that groups with highly intelligent and conscientious leaders will be well-managed and successful. Additionally, the least conscientious member of the group explained 6% of the variance in team performance. This could support the wide-spread theory that unorganized individuals, functioning as a part of an otherwise well-coordinated team, do not adopt the practices of the group, but instead drag the team down and prevent optimal performance from being achieved. Furthermore, the least agreeable member of the team explained 11% of the variance

Personality Factor	Minimum β	Maximum β	Minimum R	Maximum R	Minimum \mathbb{R}^2	Maximum \mathbb{R}^2
Extraversion	0.39	0.23	0.23	0.12	0.05	0.01
Agreeableness	0.84	0.77	0.34	0.16	0.11*	0.03
Conscientiousness	0.41	1.11	0.24	0.27	0.06^{*}	0.07^{*}
Neuroticism	-0.56	0.11	-0.23	0.06	0.05	0.00
Openness	0.10	0.33	0.04	0.10	0.00	0.01

in group performance. It seems that the "weakest link" may actually be one of the strongest predictors of team effectiveness.

*P-Value < 0.05

 Table 4.3: Individual Personality Factor Minimums and Maximums Results

Personality Facet	Minimum β	Maximum β	Minimum R	Maximum R	Minimum \mathbb{R}^2	Maximum \mathbb{R}^2
Assertiveness (E)	0.32	0.16	0.19	0.09	0.04	0.01
Activity (E)	0.25	0.28	0.19	0.15	0.04	0.02
Altruism (A)	0.61	0.07	0.29	0.02	0.27^{*}	0.00
Compliance (A)	0.35	0.27	0.20	0.10	0.04	0.01
Order (C)	0.26	0.54	0.21	0.25	0.04	0.06*
Self-Discipline (C)	0.37	0.84	0.22	0.24	0.05	0.06*
Anxiety (N)	-0.25	0.33	-0.12	0.22	0.02	0.05
Depression (N)	-0.31	-0.16	-0.15	-0.11	0.02	0.01
Aesthetics (O)	0.14	0.43	0.10	0.21	0.01	0.05
Ideas (O)	0.03	-0.25	0.01	-0.08	0.00	0.01

*P-Value < 0.05

Table 4.4: Individual Personality Facet Minimums and Maximums Results

4.4 Social Sensitivity

Finally, the effects of social sensitivity on group performance were evaluated and the results are shown in Table 4.5. Group-level social sensitivity accounted for roughly 5% of the variance in team performance, while the social sensitivities of the swarm explained 6% of the variance. This suggests that measuring social sensitivity as a swarm may better reflect the actual group dynamics of teams, thereby increasing its ability to predict team performance.

Mind in the Eyes	Group β	Swarm β	Group R	Swarm R	ΔR	Group \mathbb{R}^2	Swarm \mathbb{R}^2	ΔR^2
Social Sensitivity	-0.87	-0.90	-0.23	-0.25	-0.02	0.05	0.06*	0.01

*P-Value < 0.05

Table 4.5: Reading the Mind in the Eyes Results

4.5 Machine Learning

Prior research in machine learning has found that overly-complex feature sets often lead to cumbersome models with poor predictive capabilities. The correlation coefficient, R, was utilized to determine the top 10 features with the highest correlations to group performance. These features, listed in Table 4.6, were extracted from the data set and used to train a variety of machine learning algorithms. The performance of these algorithms were compared to the NULL model. The NULL model consistently predicts the mean target value, or the average performance across all groups in the training set, and is often established as the baseline reference for comparison. Algorithms that effectively model the underlying relationships within the data are expected to produce a lower error rate than NULL.



 Table 4.6: Features Chosen By Correlation Coefficients

Three regression algorithms were trained and validated over the data set including Multivariate Linear Regression (MLR), K-Nearest Neighbor Regression (KNN), and LASSO Regression. The LASSO Regression algorithm utilized an alternative approach to feature evaluation and selection. The features chosen by LASSO are listed in Table 4.7. Interestingly, the LASSO algorithm selected more individual minimum and maximum features, however both feature sets primarily contained factors and facets from the Extraversion, Conscientiousness, and Agreeableness dimensions of personality.

The results of the regression models are displayed in Table 4.8. As expected, the simpler models exhibited less Mean Squared Error (MSE) than the more-complex models that utilized the full feature set. Multivariate Linear Regression was not an effective algorithm for the full model, however the K-Nearest Neighbor and LASSO algorithms both produced less error than NULL. The highest performance was achieved by the



 Table 4.7: Features Chosen By LASSO

Top 10 Feature model using the Multivariate Linear Regression algorithm (0.19 MSE). This is a substantial improvement from the NULL model which generated 0.45 MSE.

Regression	NULL	MLR	KNN	LASSO
Top 10 Features MSE	0.444637	0.191778	0.323847	0.316524
Full Model MSE	0.444637	7.10079	0.367044	0.396821

 Table 4.8: Mean Squared Error for Regression Models

The regression algorithms were also compared over group-level and swam-level feature sets. As illustrated in Figure 4.2, the swarm models consistently exhibited less Mean Squared Error than the corresponding group models. This reinforces earlier findings that suggest the personality of the swarm is a better predictor of group performance than the average scores of individual team members. The lowest error was achieved by the K-Nearest Neighbors algorithm (0.62 MSE for Swarms), however none of these models outperformed Top 10 Feature MLR.



Figure 4.1: Mean Squared Error for Regression Models

Regression	NULL	MLR	KNN	LASSO
Group MSE	0.733578	1.11895	0.69526	0.833399
Swarm MSE	0.733578	1.08608	0.629615	0.704416

Table 4.9: Mean Squared Error for Regression Models

Next, groups were split into separate classes and labeled. Teams that performed in the top 75th percentile were marked as "exceptional" and the remaining teams were labeled as "average". Three classification algorithms were trained and validated including Logistic Regression, Decision Tree Classifiers, and Naive Bayes Classifiers. The classification results are listed in Table 4.10, with Logistic Regression displaying the highest accuracy score over the validation set, at 81.8%. This algorithm outperformed the NULL model, which consistently predicted "average" team performance at 72.7% accuracy.



Figure 4.2: Mean Squared Error for Regression Models

Classification	NULL	LogisticReg	DecisionTree	NaiveBayes
Accuracy	72.7273	81.8182	72.7273	72.7273
Precision	0	100	0	50
Recall	0	33.3333	0	66.6667

 Table 4.10: Accuracy for Classification Models

Finally, an artificial neural network was constructed to model the relationship between group personality composition and team performance. The network's architecture consisted of 10 input nodes, 1 hidden layer with 50 nodes, and an output layer with 1 node. Hidden nodes utilized the sigmoid activation function and the output node used a linear activation function. Dropout regularization was applied to reduce over-fitting and grid search cross validation was deployed to optimize all network parameters. The hyperparameter combinations that were tested and the optimal architecture is listed in Table 4.11



Figure 4.3: Accuracy for Classification Models

Grid Search	Parameters	Optimal
Layers	1, 3, 5, 10, 15, 20	1
Nodes/Layer	1, 10, 50, 100, 500, 1000	50
Activation Function	sigmoid, relu	sigmoid
Epochs	50, 100, 200	50
Batch Size	5, 10, 20	5
Optimizer	adam, rmsprop	adam
Dropout Rate	0.0, 0.5, 0.7, 0.9	0.5

 Table 4.11: Grid Search Cross Validation

A visualization of the network's learning curve is illustrated in Figure 4.4. As expected, the model's loss decreases over time, converging around 20 epochs. This ensures that the network is learning the true underlying relationships and developing an effective model for future inference. The optimized network was able to demonstrate 0.71 MSE against the training set and 0.76 MSE over the validation set. Regardless

of its power, the neural network approach did not achieve the accuracy exhibited by more simplistic regression and classification techniques, like Multivariate Linear Regression and Logistic Regression.



Figure 4.4: Learning Curve For Artificial Neural Network

Chapter 5

FUTURE WORK

5.1 Personality Measurement

There is great potential for future researchers to measure group personality compositions through human swarming interfaces as opposed to traditional aggregation methods like averages, minimums, and maximums. This measurement technique has been shown to produce results that better reflect the actual dynamics of the group and thereby amplify the ability to predict team performance. Further investigation into the effectiveness of these methods is advised.

5.2 Swarm Behavior

While the personality of the swarm was thoroughly analyzed, there was no investigation into the effects of swarm behavior. As swarms negotiate, they not only select different answers, but also exhibit different collective behaviors. The momentum of the puck, the factions that form, and the behaviors of individual contributors vary across swarms. Further research is advised to investigate the effects that swarming behavior has on group personality and resulting team performance.

5.3 Machine Learning

One of the biggest limitations in this study was access to data. Due to the novelty of the Swarm AI technology, administrators were required to moderate the swarms in person. Data had to be collected by hand over the course of a full school year and only about a hundred teams volunteered to participate, which is inadequate for some machine learning techniques. Ideally, hundreds or even thousands of groups would have been surveyed and the extra data would be expected to reduce model loss and improve classification accuracy.

Chapter 6

CONCLUSION

Overall, the findings presented in this paper support previous claims that group personality composition is a strong predictor of team performance. However, this study also shows that when groups of individuals gather online and form artificial swarms, they are able to amplify the predictive capabilities of their GPC by as much as 19%. Additionally, Multivariate Linear Regression and Logistic Regression were the machine learning algorithms that demonstrated the strongest performance across the board. The full benefits of human swarming is still unknown, however we can look to nature for continued guidance and inspiration. Swarms of autonomous vehicles could someday navigate traffic in ways only previously observed in flocks of birds. Swarms of legislators could someday form ideal policies and make optimal decisions only previously seen by swarms of honey bees. Finally, swarms of engineers could someday design and build structures that no individual could imagine, as demonstrated by the massive nests constructed by colonies of termites. As the human race pushes the boundaries of technology, we need to take the appropriate measures to ensure that our collective intelligences can compete with the artificial intelligences we create.

BIBLIOGRAPHY

- H. Ahmed and J. Glasgow. Swarm intelligence: Concepts, models and applications. 2012.
- [2] F. H. Al-Qahtani and S. F. Crone. Multivariate k-nearest neighbour regression for time series data - a novel algorithm for forecasting UK electricity demand. In *Neural Networks (IJCNN)*, 2013.
- [3] S. Alaliyatand, H. Yndestad, and F. Sanfilippo. Optimisation of boids swarm model based on genetic algorithm and particle swarm optimisation algorithm. In 28th European Conference on Modelling and Simulation (ECMS), 2014.
- [4] D. P. Baker and E. Salas. Analyzing team performance: In the eye of the beholder? In *Military Psychology*, 1996.
- [5] S. Baron-Cohen, S. Wheelwright, J. Hill, Y. Raste, and I. Plumb. The reading the mind in the eyes test revised version: a study with normal adults and adults with asperger syndrome or high-functioning autism. In J Child Psychol Psychiatry, 2001.
- [6] M. R. Barrick, G. L. Stewart, M. J. Neubert, and M. K. Mount. Relating member ability and personality to work-team processes and team effectiveness. In *Journal of Applied Psychology*, 1998.
- [7] B. Barry and G. L. Stewart. Composition, process, and performance in self-managed groups: The role of personality. 1997.
- [8] E. Boudreau. AI predicted oscar winners with stunning 94% accuracy. Cision, 2018.

- [9] V. Chahkandi, M. Yaghoobiand, and G. Veisi. Feature selection with chaotic hybrid artificial bee colony algorithm based on fuzzy (chabcf). In *Journal* of Soft Computing and Applications, 2013.
- [10] M. Collins. The effects of group personality composition on project team performance: Operationalizations and outcomes. 2014.
- [11] P. T. Costa and R. R. McCrae. Age-differences in personality structure: Cluster analytic approach. In *Journals of Gerontology*, 1976.
- [12] A. Cuthbertson. Artificial intelligence turns \$20 into \$11,000 in kentucky derby bet. Newsweek, 2018.
- [13] G. S. V. der Vegt, B. J. Emans, and E. V. de Vliert. Team members' affective responses to patterns of intragroup interdependence and job complexity. 2000.
- [14] M. Dorigo, M. Birattari, C. Blum, M. Clerc, T. Sttzle, and A. Winfield. Ant colony optimization and swarm intelligence. 2008.
- [15] N. R. Draper and H. Smith. Applied regression analysis. 2014.
- [16] Everwise. Why employers use personality tests. Modern Workforce, December 2015.
- [17] F. Farahnakian, T. Pahikkala, P. Liljeberg, and J. Plosila. Energy aware consolidation algorithm based on k-nearest neighbor regression for cloud data centers. In *IEEE/ACM 6th International Conference on Utility and Cloud Computing*, 2013.
- [18] A. Ferlitsch. Regression methods in machine learning simple linear regression. 2017.

- [19] V. Fonti. Feature selection using lasso. 2017.
- [20] D. Grzonka, A. Jakbik, J. Koodziej, and S. Pllana. Using a multi-agent system and artificial intelligence for monitoring and improving the cloud performance and security. In *Future Generation Computer Systems*, 2017.
- [21] T. Halfhill, E. Sundstrom, J. Lahner, W. Calderone, and T. M. Nielsen. Group personality composition and group effectiveness: An integrative review of empirical research. In *Small Group Research*, 2005.
- [22] R. Heller. Population curation in swarms: Predicting top performers. 2018.
- [23] J. Hernandez. Why birds fly in v formation. *Hub Pages*, 2015.
- [24] O. John, E. Donahue, and R. Kentle. The big five inventory-versions 4a and 54. 1991.
- [25] O. John, L. Naumann, and C. Soto. Paradigm shift to the integrative big-five trait taxonomy: History, measurement, and conceptual issues. In *Handbook* of personality: Theory and research, pages 114–158, 2008.
- [26] O. P. John and S. Srivastava. The big five trait taxonomy: History, measurement, and theoretical perspectives. In *Handbook of personality: Theory and Research*, 1999.
- [27] O. P. John and S. Srivastava. The big-five trait taxonomy: History, measurement, and theoretical perspectives. 1999.
- [28] J. Jordan. K-nearest neighbors. 2017.
- [29] J. Jordan. Logistic regression. 2017.

- [30] G. Kaur and E. N. Oberai. A review article on naive bayes classifier with various smoothing techniques. In *International Journal of Computer Science and Mobile Computing*, October 2014.
- [31] Y. Kawata, M. Hirosawa, N. Hirotsu, and E. Kirino. Relationship between empathizing-systemizing cognitive styles and mind reading among japanese adolescents. In Juntendo Medical Journal, 2014.
- [32] S. B. Kotsiantis. Decision trees: a recent overview. 2013.
- [33] O. Kramer. Dimensionality reduction by unsupervised k-nearest neighbor regression. In 10th International Conference on Machine Learning and Applications, 2011.
- [34] S. Kuiper. Introduction to multiple regression: How much is your car worth? In Journal of Statistics Education, August 2017.
- [35] M. Kulkarni. Decision trees for classification: A machine learning algorithm. The Xoriant, 2017.
- [36] F. Li1, Y. Yang, and E. P. Xing. From lasso regression to feature vector machine. 2005.
- [37] S. B. Maind and P. Wankar. Research paper on basic of artificial neural network. In International Journal on Recent and Innovation Trends in Computing and Communication, 2014.
- [38] D. Montgomery and E. Peck. Introduction to linear regression analysis. 1982.
- [39] L. Muchnik, S. Aral, and S. J. Taylor. Social influence bias: A randomized experiment. In *Science*, 2013.

- [40] A. Navada, A. N. Ansari, S. P., and B. A. Sonkamble. Overview of use of decision tree algorithms in machine learning. In *IEEE Control and System Graduate Research Colloquium*, 2011.
- [41] G. Neuman, S. Wagner, and N. Christiansen. The relationship between work-team personality composition and the job performance of teams. In *Group Organization Management*, 1999.
- [42] G. A. Neuman and J. Wright. Team effectiveness: Beyond skills and cognitive ability. In *Journal of Applied Psychology*, 1999.
- [43] S. Packianather and E. Mastrocinque. Survey paper on swarm intelligence. 2016.
- [44] B. R. Patel and K. K. Rana. A survey on decision tree algorithm for classification. 2014.
- [45] C. J. Peng, K. L. Lee, and G. M. Ingersoll. An introduction to logistic regression analysis and reporting. 2002.
- [46] B. Rammstedt, C. J. Kemper, and I. Borg. Correcting big five personality measurements for acquiescence:. In *European Journal of Personality*, 2013.
- [47] C. Reynolds. Flocks, herds, and schools: A distributed behavioral model. 1987.
- [48] I. Rish. An empirical study of the naive bayes classifier. 2001.
- [49] L. Rosenberg. How swarm intelligence could save us from the dangers of AI. Venture Beat, November 2015.
- [50] L. Rosenberg. Human swarming, a real-time method for parallel distributed intelligence. In Swarm/Human Blended Intelligence Workshop (SHBI), pages 1–7, Sept 2015.

- [51] L. Rosenberg. Human swarming and the future of collective intelligence. Singularity, July 2015.
- [52] L. Rosenberg. Human swarms, a real-time paradigm for collective intelligence. 2015.
- [53] L. Rosenberg. Artificial intelligence isn't just about intelligence, but manipulating humanity. *Futurism*, 2016.
- [54] L. Rosenberg. Artificial swarm intelligence vs human experts. In International Joint Conference on Neural Networks (IJCNN), pages 2547–2551, July 2016.
- [55] L. Rosenberg. Keeping up with AI: Why humanity must cultivate a hive mind. *Futurism*, April 2016.
- [56] L. Rosenberg. Super-intelligence and the virtues of a hive mind. Singularity, February 2016.
- [57] L. Rosenberg and N. Pescetelli. Amplifying prediction accuracy using swarm
 A.I. In *Intelligent Systems Conference (IntelliSys)*, pages 61–65, Sept 2017.
- [58] L. Rosenberg and G. Willcox. Artificial swarm intelligence vs vegas betting markets. 2018.
- [59] L. Rosenberg, G. Willcox, D. Askay, L. Metcalf, and E. Harris. Amplifying the social intelligence of teams through human swarming. 2018.
- [60] L. Rosenberg, G. Willcox, S. Halabi, M. Lungren, D. Baltaxe, and M. Lyons. Artificial swarm intelligence employed to amplify diagnostic accuracy in radiology. 2018.
- [61] B. Schneider. The people make the place. In *Personnel psychology*, 1987.

- [62] V. Sharma, S. Rai, and A. Dev. A comprehensive study of artificial neural networks. In International Journal of Advanced Research in Computer Science and Software Engineering, 2012.
- [63] D. Silver, J. Schrittwieser, and K. Simonyan. Mastering the game of go without human knowledge. *Nature*, 2017.
- [64] K. G. Smith, K. A. Smith, J. D. Olian, H. P. Sims, P. Obannon, and J. A. Scully. Top management team demography and process: The role of social integration and communication. In *Administrative Science Quarterly*, 1994.
- [65] C. J. Soto and J. J. Jackson. Five-factor model of personality. 2015.
- [66] C. J. Soto and J. P. Oliver. Ten facet scales for the big five inventory: Convergence with neo pi-r facets, self-peer agreement, and discriminant validity. 2008.
- [67] R. Tibshirani. Regression shrinkage and selection via the lasso. In Journal of the Royal Statistical Society, 1996.
- [68] F. Walumbwa, B. Avolio, and W. Zhu. How transformational leadership weaves its influence on individual job performance. In *Personnel Psychology*, 2008.
- [69] G. Wiederhold and J. McCarthy. Arthur samuel: Pioneer in machine learning. In IBM Journal of Research and Development, 1992.
- [70] A. W. Woolley. Evidence for a collective intelligence factor in the performance of human groups. In American Association for the Advancement of Science, 2010.