2020 International Conference on Intelligent Data Science Technologies and Applications (IDSTA)

A Comparison of Data-Driven and Traditional Approaches to Employee Performance Assessment

Nirvan Sharma, Patrick Hosein
The University of the West Indies
Trinidad and Tobago
Email: nirvan.sharma@my.uwi.edu, patrick.hosein@sta.uwi.edu

***Abstract*—Employee performance monitoring is not new. Traditional systems consider an average of peer and direct supervisor reviews across various subjective criteria in yearly performance appraisals. This leads to problems such as recency and social biases that devalue the credibility of the review process. Modern, data-driven systems have made it easier to automatically track employee data regarding various common performance metrics in useful ways. Statistical measures are derived from these metrics that are collated over a defined period. These measures are used to determine if and when any appropriate action needs to be taken. However, the statistics fail to account for common characteristics of the employee population. Instead of comparing individual performance against rigid idealistic norms, this paper presents an automated and objective approach to determine the collective population norm and calculate abnormal deviations from it. Employees who deviate from the norm are automatically flagged and subject to further review from management. Ideally, a combination of calculated metrics using this approach can be used to accurately reflect real-world situations and highlight anomalies in the employee work force. We compare the results of the proposed statistical approach with traditional human review approaches and evaluate their correlation.**

***Index Terms*—Anomaly Detection, Data Analytics, Business Intelligence, Performance Review**

I. Introduction

The productivity of a company is directly related to the performance of its employees. As an employee, performance reviews are a way to prove your worth and strive for promotions or higher wages. As a company, reviews are important to track employees' abilities in order to effectively leverage their skills. Employee dissatisfaction can occur due to an unfair performance review and reward system. Employees become discouraged if their efforts are not recognized[1]. Studies show that there is a correlation between average job satisfaction and productivity[2]. Companies also suffer from a sub-optimal system as employees can find ways to manipulate results. Large companies commonly track various employee metrics already, for example, attendance and time worked. They use aggregations of these metrics to determine employee average performance. Data-driven systems like this are more objective and accurate compared to older performance review strategies such as 360 degree feedback loops. However, note that individual performance evaluation for current data-driven systems are also determined independently of the individual's coworkers and their performances. External factors can contribute to an employee's attendance, performance, and productivity. These same factors can also affect the performances of many other employees. In order to be fairer in review, similar employees are manually compared for that period, and a decision is made if similar performance was recorded for the examined employees. This determines whether or not an employee's performance is justified. Consider the following scenario where a company chooses to distinguish workers who live very far from the office and workers who live close by. It is likely that the long distance group often gets caught in traffic more than the closer group and consequently has a higher variance in clock-in times. A fair approach would be to be more lenient on the latter group as they would find it harder to arrive earlier than the former group. In this paper, we propose a more meaningful approach to statistical performance monitoring whereby we compare the performance of the employee's peers with their own performance to determine anomalous behaviour. In this paper, we start of by discussing the related solutions and contributions around the existing problems and specify how the proposed approach can address some of the problems that were found (Section 2). Next, we briefly describe the datasets at our disposal and how we will be using the data (Section 3). We follow with our generalized proposed approach that would fairly evaluate employee performance and flag performances that deviate from the norm (Section 4). We then show examples of how a company can implement our approach to track different employee performance metrics and combine these calculated metrics into one output for management to review. In Section 5, we investigate ways in which we can compare our results versus the current review system to estimate the accuracy of both approaches. In summary, we demonstrate an automatic solution that detects abnormal employee performances and we try to determine it's effectiveness alongside an existing traditional method.

II. Related Work and Contributions

Employee satisfaction is a key objective for businesses. This idea is supported by psychology researchers such as Gregory[1] and Greene[3], who emphasize that employee satisfaction is essential to the success of a business as it directly relates to performance and productivity. Studies show that a predictor of job dissatisfaction is burnout[4]. Burnout correlates to increased risk of future illnesses and it is implied that preventing burnout will reduce the rate of absenteeism and tardiness. Absenteeism and tardiness also correlate with

978-1-7281-8376-3/20/$31.00 ©2020 IEEE

an employee's perception of averse work conditions[5]. If a company has a clear review system with target performance goals that can translate to promotions and remuneration then employees have more incentive to perform. If the review system could automatically track predictors of burnout and employee dissatisfaction, the benefit is two-fold.

Research has been done to tailor the existing manual appraisal process to be more accurate and objective. One of the main disadvantages regarding a traditional system is that performance criteria is prone to favouritism or prejudice, and often wrongly considers only the small period before the review itself. Islama et. al.[6] created an AHP-based (Analytic Hierarchy Process) evaluation process based on weighted criteria to combat such problems. The criteria was structured around quantity/quality of work, planning/organization, teamwork/cooperation and more, weighted by importance by the Human Resource Managers. Each employee was given a rating on their performance on each weighted criteria and an overall weighting score was calculated. The research suggested strict guidelines to be followed in a revamped appraisal workflow to encourage a consistent, fair review system. However, this system remains manual and is heavily reliant on the Human Resource Department's willingness to cooperate.

To address this, the first step we take is to create a system that is automatic and accurate. A time and attendance system was devised in a study[7], to promote accurate labor reporting. This was a time card system with a card reader that recorded employees' entries and exits. The difference in hours worked by an employee as well as their minutes was calculated and output as reports. Note that evaluations for employees were done solely on their data and compared to a company baseline figure.

Several solutions, example[8] and[9], make use of Internet Of Things (IoT) based systems to automatically gather accurate data that feeds into an evaluation algorithm. A limitation of our proposed solution is that it relies on the accuracy and integrity of the data provided. The attendance dataset used in this paper was derived from Radio-frequency identification (RFID) scanners for recording clock-in and clock-out times. While IoT and RFID devices do not eliminate all methods of data tampering, they discourage a variety of them.

Kaur et al[9] classified their IoT data into three implications (Positive, Negative, Neutral). They consider factors such as employee location and activity to calculate employee implication based on some cognitive decision making using fuzzy logic. Industrial activities such as Water Consumption, Material Theft and Team Work were monitored and the employees' actions were graded. They found that employees performed better when they knew that they were being monitored by an external system that rewards Positive actions. Similar results were found by Sood et. al.[8] who implemented a system to effectively motivate employees as well as evaluate their performance. In their system they used a sophisticated game theory model to evaluate employee implications instead of fuzzy logic.

Our proposed solution builds on top of the acknowledgement that a performance review system must be automatic and data-driven. What we contribute is a method that can automatically and more fairly evaluate employees based on the data surrounding their peers' performances. There is an implication that employee performance is nuanced around many social and environmental factors that can be captured if such factors affect the employee population's performance as a whole. We provide a tool to save employers time and resources as they no longer need to individually review persons but only those who are flagged by the system.

III. Description of Datasets

The datasets consist of distinct attendance and overtime records spanning 3 years on 100 shift and 100 non-shift workers at a large factory. Employee clock-in and clock-out times have been recorded daily and cross referenced with overtime records to determine total work hours for that day. Only supervisor/manager approved overtimes were referenced, not unapproved overtime requests. The combined dataset is similar to Table I, which is sorted by attendance date. Each row contains the attendance date that was recorded, the total hours worked by an employee, the identification number of the employee in attendance and whether or not the person worked overtime on that date. We can then account for any hours worked overtime for an employee given that a baseline 8.5 hours is required daily. The datasets were split by shift and non-shift employees because only shift workers were primarily subject to performance reviews. However, our proposed approach works the same when accounting for both shift and non-shift workers. We use these datasets to provide us with information around different employee statistics to determine the population norms of workers over the years and as input into our proposed approach. Additionally, we have a dataset of performance appraisal ratings that were given to each employee by their supervisor/manager. We will use this dataset to compare our results and to evaluate the effectiveness of each performance review method. We were not given permission to publicly share the datasets as they contain private information about the factory's workers and their whereabouts.

IV. Proposed Approach

We assume that various metrics (e.g, hours worked, overtime) are recorded for each employee on a periodic basis (e.g, daily). We also assume that the data of the employee population for each tracked metric can generally be described by a Gaussian distribution. Given normal working conditions over a significant period of time, this would be a safe assumption to make on the distribution of collected data. A company may choose a bi-annual or annual reporting period where they would like to evaluate employee performance. Let T be one such reporting period where data is collected at some periodic frequency, t. One of the standard metrics recorded for an employee can then be denoted by $x_i(t)$. The sample metric mean for employee i is:

$$\bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_i(t) \tag{1}$$

Let N be the number of employees being recorded. Due to our prior assumptions of this parametric model, if N is small, T must be sufficiently large or vice versa in order to collect meaningful results. Given that each employee's sample mean has been calculated, we can then determine the sample population mean and standard deviation denoted by:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{2}$$

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(\bar{x}_i - \bar{x})^2} \tag{3}$$

Typically an ideal value for a given metric (e.g, 8 hours worked per day) can be denoted k_x. We use Grubb's test (also known as maximum normal residual test) to determine the performance measure of an employee. We use these calculated metrics to detect anomalies in the population. To increase fairness further, k_x is taken as the population mean. The result is that each employee is directly and fairly compared to the population norm. Let X_i be the normalized variation of the employee recorded metric.

$$X_i = \frac{|x_i - k_x|}{s} \tag{4}$$

X_i is how far an employee's sample mean is from the sample population mean in units of sample standard deviations. An employee is said to be anomalous if their metric exceeds 2. According to the Gaussian Distribution empirical rule, the probability of being greater than 2 standard deviations away from the mean is less than 5%. We believe that this is a good basis for flagging an employee as anomalous especially if N is large.

This can be repeated for multiple metrics and a final performance metric can be derived from a weighted average of all metrics. The weights of each calculated metric depends on the importance of the metric. For example, time worked can be weighted more than minutes late. Lastly, the final calculated performance metrics can be recomputed for every reporting period and performance can be tracked over time.

The company currently has an annual reporting period where they run a 360 degree feedback review process. One of the factors in evaluating an employee is called Dependability. A Dependability rating is usually given to an employee by his peers and manager. In this case, dependability is an important consideration for shift workers as the company would prefer to always have one person at a station at all times in a factory. A person should not freely leave their post until a worker from the upcoming shift comes to relieve them. We can relate dependability by applying our approach to a combination of

TABLE I
SAMPLE ATTENDANCE RECORDS FOR COMBINED DATASET

ID	Date	Overtime	Total Hours
0	13-May-2019	No	8.5
1	13-May-2019	Yes	10.5
2	13-May-2019	No	8.5
0	14-May-2019	Yes	12
1	14-May-2019	Yes	10.5

recorded metrics such as work hours and attendance. We also apply our algorithm to non-shift workers to compare performance between the two.

A. *Time Worked Example*

Consider an example of tracking hours worked daily in a large factory with both shift and non-shift workers. The company would like for an employee to work 8.5 hours given that the employee stays on the compound for their 30 minute lunch break. The first metric we consider is the average number of hours worked daily by an employee. If the population on average works 8.5 hours a day, then employees who consistently work 8.5 hours daily should automatically be given a low metric rating while employees who work significantly more or less than that should be flagged. An employee with a metric value of 0 would be the ideal employee. We used the proposed approach to compute the metrics for approximately 200 employees in our dataset and found interesting results.

1) Shift Workers: While the company regulates an 8.5 hour work day, the average shift worker works closer to 9.5 hours a day. We believe this can be due to the frequency of overtime approved. Overtime of at least 2 hours is approved for shift workers who are not relieved from their post after 1 hour of scheduled shift end. Thus, it is common to find employees working more than they have been scheduled to work. However, no person should work more than 16 hours in a row. In this scenario, given that the standard deviation of the sample population was found to be approximately 0.6, an employee who works less than 8.5 hours daily would be anomalous for this considered metric. This scenario may seem unfair from an employee's perspective as that person is working the required amount of hours. However, from a company's perspective, flagging this employee can still be important as they statistically work less than average. The system will simply flag persons who deviate from the norm and thus it is up to management to decide further action. Plotting a histogram with 8 bins, seen in Figure 1, describes the distribution of the shift workers' daily time worked. If the shape of the distribution is not Gaussian, the algorithm provides similar results. An example of this would be for the distribution of non-shift workers and their average daily work hours, Figure 4.

For each reporting period, we can show the performance of the sample population on a bar graph at a more granular level

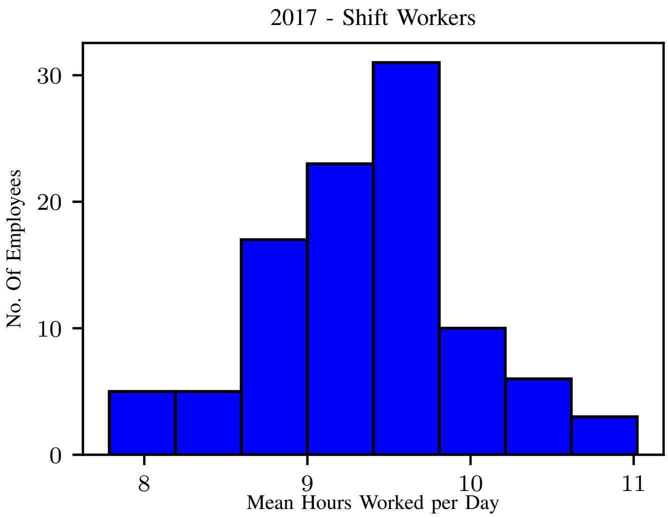

Fig. 1. Shift workers' average daily working hours in 2017

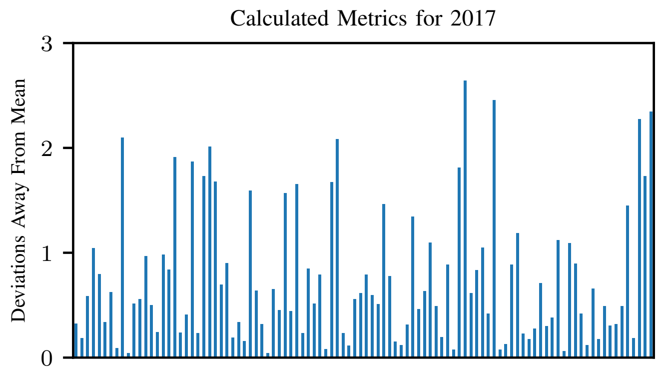

Fig. 2. Employees' metrics for a single reporting period (2017). Each bar represents an employee's deviation. The taller the bar, the more anomalous the person's performance.

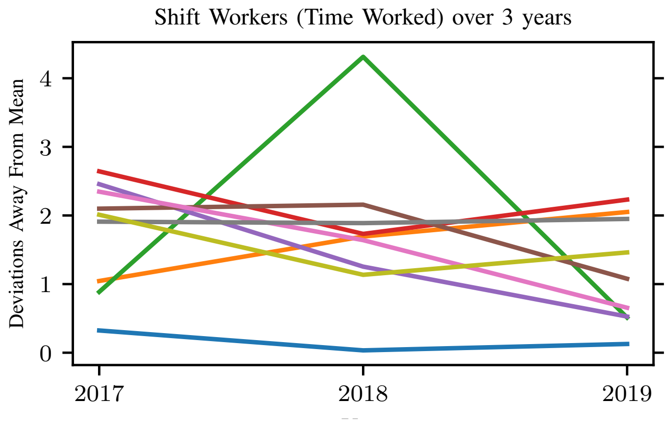

Fig. 3. Tracking employee performance(in terms of hours worked) over 3 years. Each line represents an employee's deviation over time.

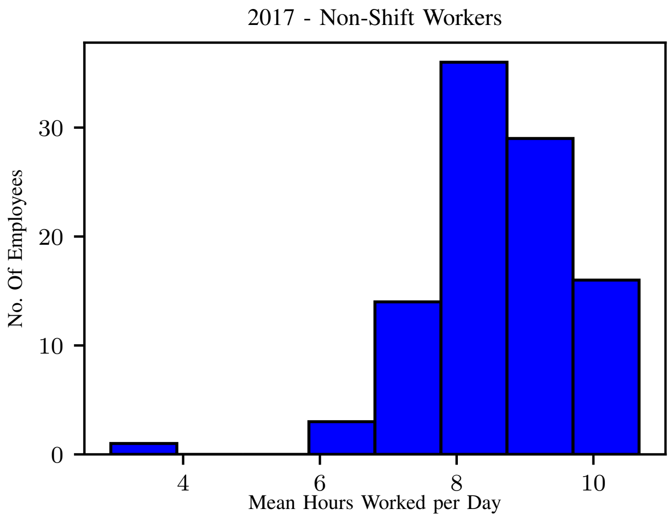

Fig. 4. Non-shift workers average hours worked daily.

(Figure 2). Each bar represents an employee for that reporting period and the height of the graph represents their calculated deviations.

The sample results of the calculated time worked metric over 3 years can be found in Figure 3. Each employee is represented by a plot on the graph which tracks their performance over the reporting period. For more detailed tracking, this method can be done monthly and output for each month instead of year. From the results, an employee who is greater than 2 standard deviations away from the mean works approximately 1 hour and 20 minutes more, or less, than an average worker.

2) Non-Shift Workers: Overtime is not as common for non-shift workers and, as a result, the mean time worked daily is approximately 8.6 hours for an average non-shift employee. Therefore in 2017, shift workers worked consistently more hours than their non-shift counterparts. In a traditional performance review system that does not consider the population's features, this difference in time worked could potentially be misleading if shifts and non-shifts reviews were not separated. In this case, the company only uses 360 degree feedback reviews on their shift workers as mentioned before. Figure 4 is a result of plotting a histogram of the average time worked daily by non-shift employees. From the histogram, if an individual is working as low as an average 3-4 hours, our system will automatically flag this person for review. Then management can further determine the reasoning for such an outlier. In this case, the employee frequently had work meetings outside of the compound.

B. Attendance Example

Attendance is an appropriate metric to consider pairing with time worked because it gives more context into an employee's performance. More context is necessary because it is possible for an employee to affect the company's productivity by taking too many unapproved leaves but ensuring they work enough hours when they do attend work. We resolve this issue by

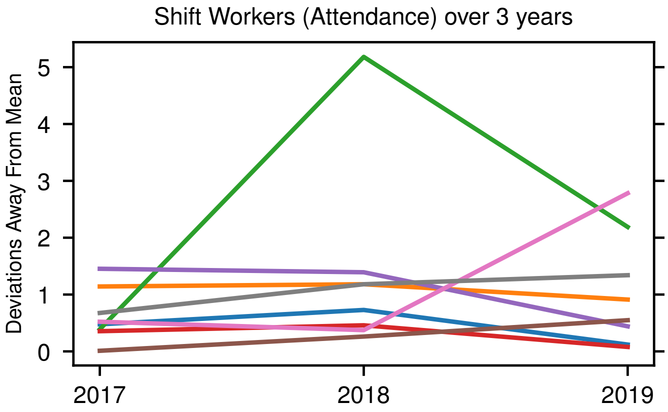

Fig. 5. Tracking employees based on an attendance-based metric over 3 years. Each line represents an employee's deviation over time.

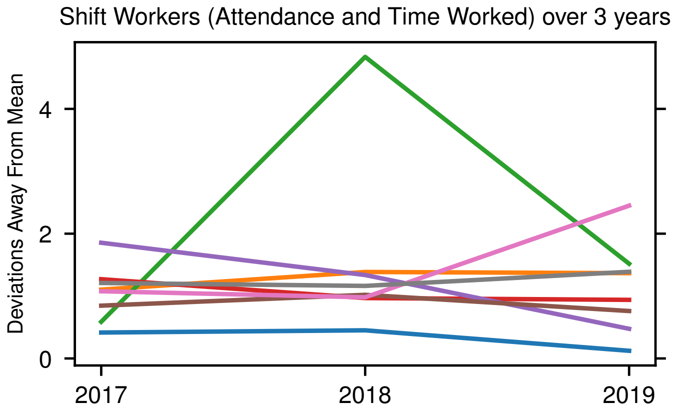

Fig. 6. Results of combining two metrics with a weighted average. Each line represents an employee's deviation over time.

considering employee attendance as a metric. We measure each employee's attendance percentage by using the ratio of number of days attended over possible working days for that employee for the reporting period. Employees who have a very high (85% or more) or very low (40% or less) attendance percentage will ideally be flagged as anomalous. Average shift and non-shift workers work approximately 210 and 208 days respectively for the year with an overall greater than 80% attendance. The sample results of plotting the attendance metric over 3 years for the same shift employees can be seen in Figure 5.

C. Combining Metrics

When combining multiple data-driven performance metrics we would like to preserve the anomalous nature that may be found in each calculated metric. We achieve this using the weighted average of the chosen metrics. In this scenario we will be looking at time worked and attendance as our metrics. The overall influence an individual metric has on the final output of the algorithm depends on the weighting given to the metric. It is rare for employees to be anomalous in several aspects and so we encourage tracking of multiple metrics to determine overall productivity. This places more emphasis on the employees who get flagged by the system as they would be anomalous across multiple important metrics and certainly require attention. In general, the effectiveness of the proposed approach increases as more metrics are tracked. These combined metrics would give a holistic summary of employee performance. Examples of other metrics that can be tracked are (a) Minutes Late, (b) Company-specific key performance indicators, for example, station inspections, products sold or units manufactured and (c) Unapproved Leaves taken.

In our weighted average calculation in this experiment, Time Worked and Attendance were given weights of 0.4 and 0.6 respectively. We place slightly more importance on attendance as it is a bigger factor in being "dependable". Given the datasets, this was the closest estimate of the dependability attribute inside the current traditional appraisal process. Figure 6 is the output of running a weighted average of the two metrics over the 3 years. Attendance is weighted slightly higher than hours worked and the outcome of that is shown on the graph, where the trajectory of the employee's lines have been more influenced by their performance in terms of attendance.

V. Traditional Approach

The traditional review method consists of common human resource strategies like 360 degree feedback reviews or performance appraisal systems. A typical appraisal begins with a single review by a peer, supervisor or the review subject themselves. An example of the current system can be seen in Table II. In this example table, a supervisor has manually assessed an employee on several subjective criteria and assigned a score to each. The scores for each feature, ranging from 1-5 are compiled into one value using a weighted average based on company preference. An employee's peers and supervisor/manager will each rate an employee as shown in the example and a weighted average of all reviews for an employee would be taken where a supervisor/manager's score will be given the highest weight. This final output is the overall rating for an employee for the performance period.

Lunenberg[10] discusses the idea of 'rating errors'. Rating errors are psychological tendencies and biases that can affect the accuracy of performance appraisals from supervisors. An example of a rating error is called Central Tendency. Supervisors may be reluctant to rate subordinates very high or very low. They prefer to not be considered too strict by giving a low rating and conversely, they would rarely feel that a subordinate deserves the highest rating. This leads to everyone having an average rating. For example, looking at the distribution in Figure 7, we see that the rating scale is heavily centered around scores of 3 to 4. However, there are some people who scored higher than 4 who seem to be positive outliers. Our proposed data-driven approach is different because it is able to detect both positive and negative outliers in an automatic and objective manner. We would like to evaluate the output

TABLE II
SAMPLE PERFORMANCE REVIEW OF AN EMPLOYEE

Performance Criteria	**Score(1-5)**
Job Knowledge & Comprehension	4
Skill	3.4
Productivity	4
Leadership	3.7
Management Principles	3.8
Dependability	4
Initiative/Problem Solving	3.9
Interpersonal Skills	4
Safety	3.5
Calculated Weighting	**3.86**

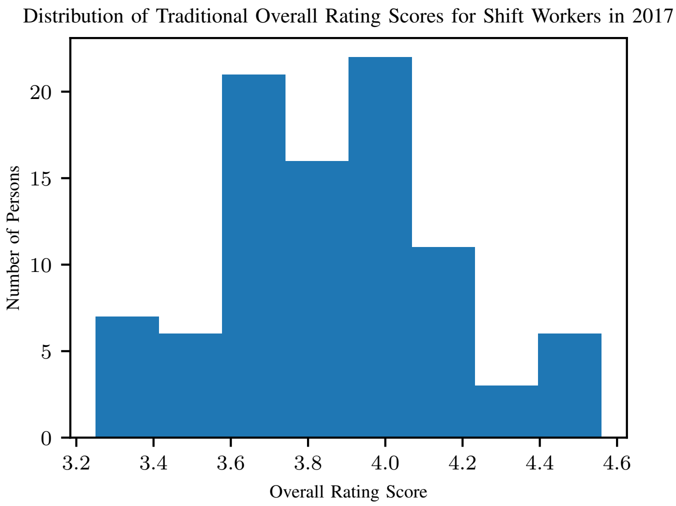

Fig. 7. The distributions of normalized features for the traditional approach

from both the existing traditional approach and the proposed data-driven approach for the year 2017 to determine if both methods arrive at similar conclusions.

VI. Comparison

When comparing the two methods, we especially want to determine the ability of both assessments to highlight anomalies in the employee population. Ideally we would want to see a correlation or agreement between both methods on existing outliers. However, we found that this was rarely the case.

1) Correlation: We established that different errors and biases exist in the current subjective appraisal system. Therefore we cannot confidently consider it's results as the ground truth in order to validate the results of our anomaly detection approach. However, we want to find any possible correlations between the two. We can run our anomaly detection algorithm on the current system's final outputs for each employee. Before we compare both systems we use mean normalization on both output features so that they can approximately be the same

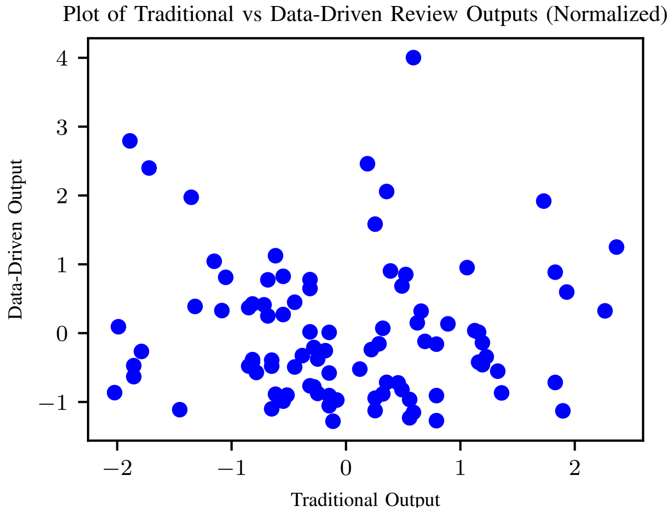

Fig. 8. Plotting the normalized traditional appraisal system's output versus the normalized data-driven output.

scale. It's important that we do not use min-max normalization as this method tends to not retain the numerical importance of our outliers. For each normalized employee overall rating, we can also find the absolute difference, or the delta, between the rating and the normalized anomaly detection output. The ideal delta between each comparison would be 0, that is, there is no difference between the data-driven approach used and the current subjective review system. We plot our results of the comparison in Figure 8.

We can observe from the scatter plot that no clear correlation exists between the results of the traditional approach and the data-driven approach. This lack of correlation is a point of concern that we would like to emphasize because it means that the traditional approach's output is not supported by data surrounding employees' performances. Calculating the correlation between the two normalized datasets using the Pearson correlation coefficient gives a result of -0.0342. A Pearson coefficient that is close to zero indicates that there is no linear relationship and at best, a very weak negative relationship. The average delta between the two approaches is 1.137 which means that there is not often a consensus between an employee's performance.

2) The analysis of differences: Correlation determines whether a relationship exists between one method or the other. However, finding a correlation or the lack of correlation can be misleading if we want to compare the agreement of our traditional and proposed approaches and not just the strength of their relationship. What we would also like to do is quantify any differences and biases between the two. For this we use the Bland Altman (B&A) plot, Figure 9, with our normalized datasets. We cannot use a B&A plot to determine if one method is preferred over the other but rather we can use it to show how the difference between the two outputs varies from some *true* measurement. In the absence of that true measurement, we use the average of the two

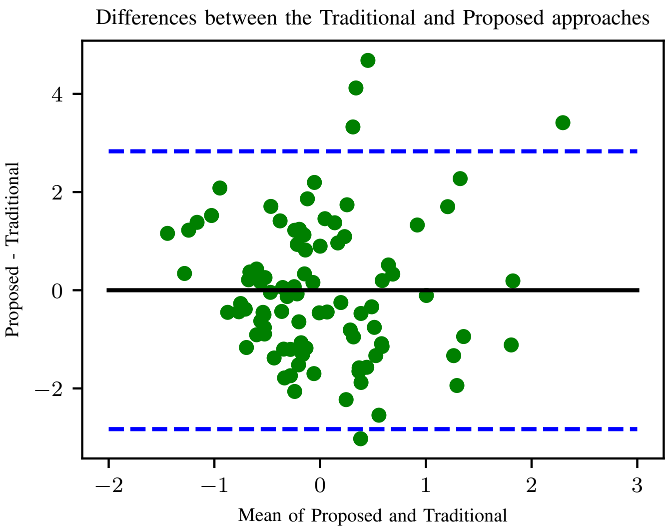

Fig. 9. B&A plot of the differences between Traditional and Proposed outputs.

outputs. The results of the B&A method can be seen in Figure 9. The difference computed is the result of subtracting the proposed approach to the output of the traditional. Therefore, any points above 0 on the plot indicate that the proposed approach gave a higher score than the traditional and vice versa. The blue lines illustrate the limits of agreement which are the measure of the spreading of differences as well as the random error of measurement. What this means is that there is a 95% probability that a new review from both methods should fall between those ranges of differences. We see that both methods are generally in disagreement, besides average reviews around the 0 mark on the x axis. There is large variability in the differences between both outputs, where the proposed approach seems to have the bigger difference for employees with potentially anomalous performances. This confirms that there is neither any significant correlation or agreement between the two methods.

3) Examples: It is difficult for reviewers to be on the same page and have access to the same data when reviewing employees. This is evident from the variance in overall ratings of employees in the appraisal process. Table III illustrates the top 5 cases where both performance review systems differ the most, ranked highest to lowest. These are taken from the shift worker dataset with mean 9.5 hours worked daily and mean 80% attendance rate. Keep in mind that the higher the data-driven or proposed approach's score is, the more anomalous the person is and that the scale for traditional ratings are from 1-5. There exists inconsistencies where persons like employee A, who have been anomalous in both metrics, have been assigned a 3.29 rating which is numerically average. At the same time, others who have not been anomalous in any regard also get similar average rating scores. Employee E, with 82% attendance rate and an average performance was given a rating greater than 4 while Employee C with only 42% attendance also got a rating greater than 4. It seems

TABLE III
EMPLOYEES WITH THE HIGHEST DELTA (AFTER NORMALIZATION) BETWEEN TRADITIONAL AND DATA-DRIVEN OUTPUT

Empl.	Delta	Time(Hrs)	Attend.(%)	Proposed	Traditional
A	4.68	7.90	59.1	2.01	3.29
B	4.12	10.67	61.3	1.82	3.34
C	3.41	8.35	42.5	2.60	4.03
D	3.33	9.62	52.5	1.62	3.45
E	3.03	9.40	82.0	0.12	4.42

that a reviewer may place higher value on subjective criteria as opposed to the objective metrics tracked. One interesting example inside the traditional system would be to look at Employee B who worked an average 10.67 hours daily and has a 60% attendance rate. One may think that the overall rating for that person should automatically be more than a mediocre 3.34. If attendance was that important for it to be B's downfall, Employee C would not have gotten a 4.03 overall with their 40% attendance. Any inconsistency in rating like what employee B received in the example would cause frustration and lead to burnout and fatigue. This also disillusions the employee's belief in the accuracy of the review process. Subjective criteria can be invaluable, but a focus on subjective criteria leads to higher variance and a certain amount of unfairness as reviews become harder to analytically measure and compare between employees. In this case, the workers are unionized and have more leniency with their performance. Because of this, the appraisal process may focus on more subjective criteria since it would be harder to penalize the performance of an employee without the union protecting them from any repercussions. Using a more data-driven approach that tracks several metrics, it may be easier to convince the union of an employee's misconduct.

If we look at the distribution of the final calculated metric scores derived from our proposed approach in Figure 10, we see that it is positively skewed, which in this context is ideal. Assuming the data is accurate, we want the majority of the employees to have a low rating that is close to 0. We also want to be able to detect outliers confidently which can also be seen at the tail end of the plot. In comparison, the distribution of scores from the traditional method, shown in Figure 7, does not use the full scale from 1-5. Therefore making it difficult to separate definite outliers to the rest of the population. An accurate and insightful review system should be able to score employees from best to worst with confidence and justification.

VII. Summary and Conclusions

We discussed the importance of employee satisfaction in industry as it relates to productivity and profit for a company. One way to motivate employees is to have clear incentives and room for individual growth in the business. Thus, having a reliable and objective way to track and evaluate employee performance is critical.

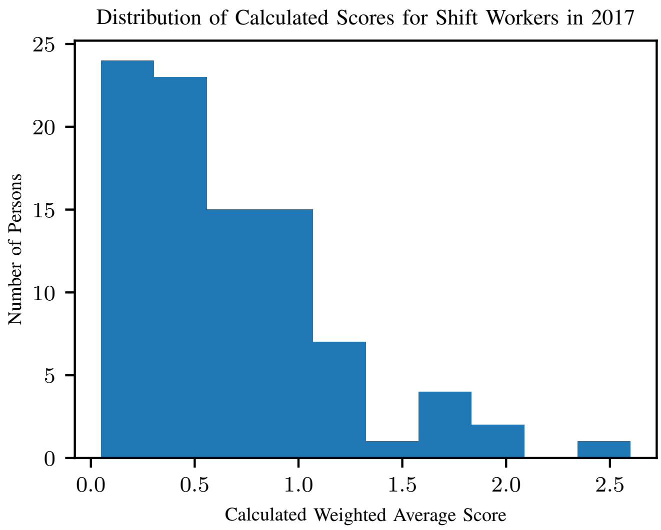

Fig. 10. The distributions of normalized features for proposed and current approaches

We proposed an approach that will automatically track employees fairly on various common performance metrics. The approach more importantly flags anomalous employee performance. Anomalous employees can be positive, i.e the employee exceeds average performances, or negative where the employee under-performs. In the negative case, a flagged individual often means that action from management is required to investigate problems such as potential or existing burnout and stress. In the positive case, a flagged employee should be considered for rewards and bonuses. This allows the company to support their employees more significantly which leads to greater employee satisfaction and higher productivity. We tried to define and compare an existing traditional review process to the proposed data-driven approach and found that no real correlation or agreement existed between two outputs. We established that feedback reviews often were not aligned with the relevant employee performance data and were generally inconsistent. Data-driven approaches like the one proposed is objective and directly supported by the data that exists. This data can be referenced and used as visual aids during in-person performance review meetings. The proposed approach importantly offers a level of automation and efficiency that is currently not possible with existing traditional systems.

References

[1] K. Gregory, "The importance of employee satisfaction," *The Journal of the Division of Business & Information Management*, vol. 5, pp. 29–37, 2011.

[2] S. Appelbaum, J. Adam, N. Javeri, M. Lessard, J.-P. Lion, M. Simard, and S. Sorbo, "A case study analysis of the impact of satisfaction and organizational citizenship on productivity," *Management Research News*, vol. 28, pp. 1–26, 05 2005.

[3] C. N. Greene, "The satisfaction-performance controversy: New developments and their implications," *Business Horizons*, vol. 15, no. 5, pp. 31–41, 1972.

[4] S. Toppinen-Tanner, A. Ojajärvi, A. Väänaänen, R. Kalimo, and P. Jäppinen, "Burnout as a predictor of medically certified sick-leave absences and their diagnosed causes," *Behavioral medicine*, vol. 31, no. 1, pp. 18–32, 2005.

[5] M. Biron and P. Bamberger, "Aversive workplace conditions and absenteeism: Taking referent group norms and supervisor support into account." *Journal of Applied Psychology*, vol. 97, no. 4, p. 901, 2012.

[6] R. Islam and S. Rasad, "Employee performance evaluation by the ahp: A case study," *Asia Pacific Management Review*, vol. 11, pp. 163–176, 01 2006.

[7] B. T. Jedrzejewski, Tomasz and A. Zgrzywa, "The analysis of data collected by time and attendance systems," *Third National Conference on Scientific Data Processing Technologies*, pp. 83–90, 2010.

[8] N. Kaur and S. K. Sood, "A game theoretic approach for an iot-based automated employee performance evaluation," *IEEE Systems Journal*, vol. 11, no. 3, pp. 1385–1394, 2017.

[9] J. Kaur and K. Kaur, "A fuzzy approach for an iot-based automated employee performance appraisal," *Computers, Materials and Continua*, vol. 53, pp. 24–38, 01 2017.

[10] F. C. Lunenburg, "Performance appraisal: Methods and rating errors," *International journal of scholarly academic intellectual diversity*, vol. 14, no. 1, pp. 1–9, 2012.