Intex II – Fag el-Gamous

04.14.2023

—

Group 4-3

Andrew Alley

Jon Stauffer

Joshua Allen

Mitch Anderson

Project URLS

Website URL	https://www.thewayofthebuffalo.de	
.ipynb URL	https://colab.research.google.com/drive/1FGJ6WkrXz2IUw2zk2GfOfzk803_6SHtt?usp=sharing **GetPass Password:** admin123	
Presentation Slides	https://docs.google.com/presentation/d/1W9DEry_TFEATDFPcdrvVp9hWgiMH2im6UVaJPHTOAII/edit?usp=sharing	
Repository FrontEnd URL	https://github.com/staufjonbyu/4-3intex-buffalo.git	**GitHub Branch:** master
Repository BackEnd URL	https://github.com/andrewmalley725/intex-2023-api	**GitHub Branch**: main

Project Credentials

Username	Password	Role
mitch@gmail.net	ThisIsMyPassword123!	Researcher
stauffjon@gmail.com	admin123	Admin (Only account that allows a weaker password)

Solution Details by Class

IS 413

- Used React frontend with ASP.NET API backend
- Utilized ONNX files to deploy supervised models
- All listed requirements should be found easily by navigating the website

IS 404

- See “AWS Services Used” and “Recommended Services” below
- Created a CNAME alias at thewayofthebuffalo.de
- Cost Estimate:
    - Assuming the site will get ~100 monthly users and all costs will be paid monthly, we estimate that our site will cost approximately $293.18 per month to run on AWS. This estimate accounts for $5.33 for a single t4g.small EC2 instance and $287.85 to run our database on RDS

IS 455

Supervised Models

- Wrapping Regression
    - Used data from Burial Main and Body Analysis Tables
- Wrapping Decision Tree
    - Used data from Burial Main and all Textile tables
- Head Direction Decision Tree
    - Used data from Burial Main and all Textile tables
- Sex Prediction Decision Tree
    - Used just data from Burial Main table
- Sex Prediction w/Textiles Decision Tree
    - Used data from Burial Main and all Textile tables

- All decision trees were trained and then re-trained using only the features that made a measurable impact on the model
- Data was cleaned and instantiated using our best judgment while trying to maintain the integrity of the data
- All decision tree models were deployed using ONNX files
- Our website allows users to input feature values and get a prediction in return
- Findings described in executive summary

Unsupervised Model

- We used K-Modes Clustering to identify trends in the data

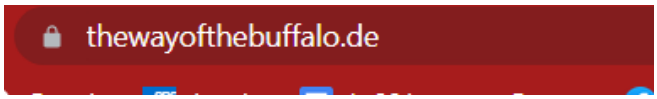
- Used data from Burial Main and all Textile tables
- 2D and 3D visualizations provided in Colab file and deployed on our site
- Findings described in executive summary and on site

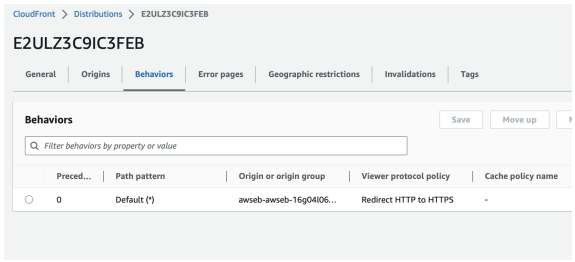
IS 414

Below are descriptions or screenshots demonstrating how we fulfilled each requirement

I. Encryption

We achieved Encryption through multiple methods, including aws cloudfront where we achieved a TLS certificate as well and an HTTPS certificate. Beyond that our Api passed through cloudfront to our front end AWS Amplify all received the proper HTTPS encryption needed. We also used AWS CloudFront to redirect HTTP traffic to HTTPS

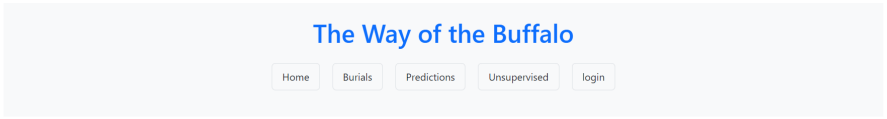


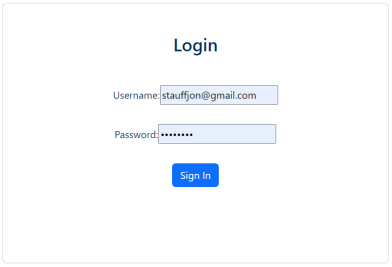


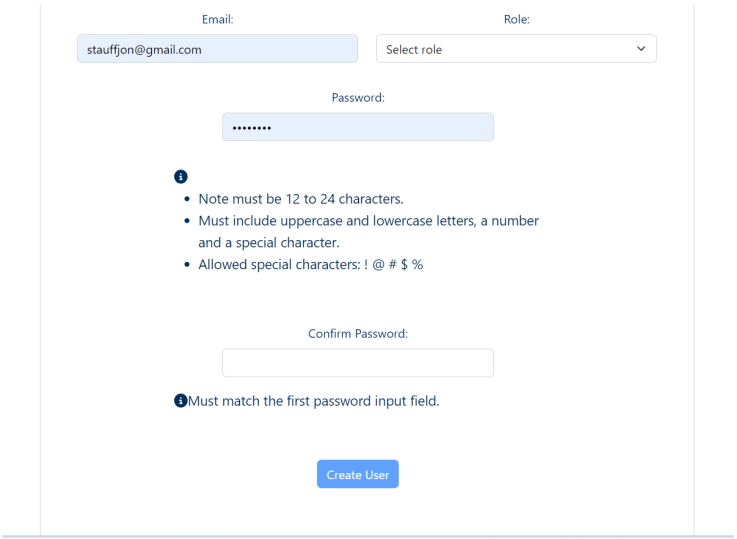
II. Authentication

User Sign In -- We implemented our own sha256 hashing to store the password on the back end! We also created our own 2fa that emails a code to the user email and requires verification.

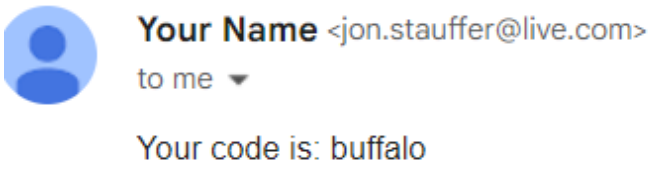
Better Password -- We invite you to try and create a password less than 14 characters. Our create user prevents all but the best passwords







Two-Factor-- our two factor was not complete however we had this code that correctly sent a verification code to the user. The problem was beanstalk blocking SMTP protocols when deployed.



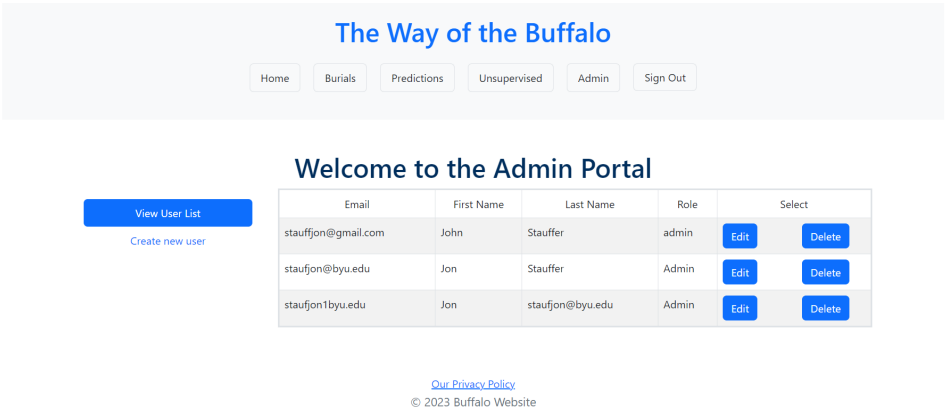

```
private void SendEmail(string email, string body)
{
    var message = new MimeMessage();
    message.From.Add(new MailboxAddress("Your Name", "jon.stauffer@live.com"));
    message.To.Add(new MailboxAddress("", email));
    message.Subject = "verification code";

    message.Body = new TextPart("plain")
    {
        Text = body
    };

    using (var client = new SmtpClient())
    {
        client.Connect("smtp.outlook.com", 587, false);
        client.Authenticate("jon.stauffer@live.com", "number20");
        client.Send(message);
        client.Disconnect(true);
    }
}
```

III. Authorization

Our RBAC is completely functional and users cannot add, edit, or delete unless they are signed in.

IV. Integrity

The Admin page is only accessed if the user is signed in and the burial page changes to allow edit if the admin is logged in.



V. Credentials

This code was added to store the hash of each password with a pepper so the user code is new stored in our database. It then authenticates the user by hashing the password and only authenticates if the password is correct. Moreover, the api only receives a

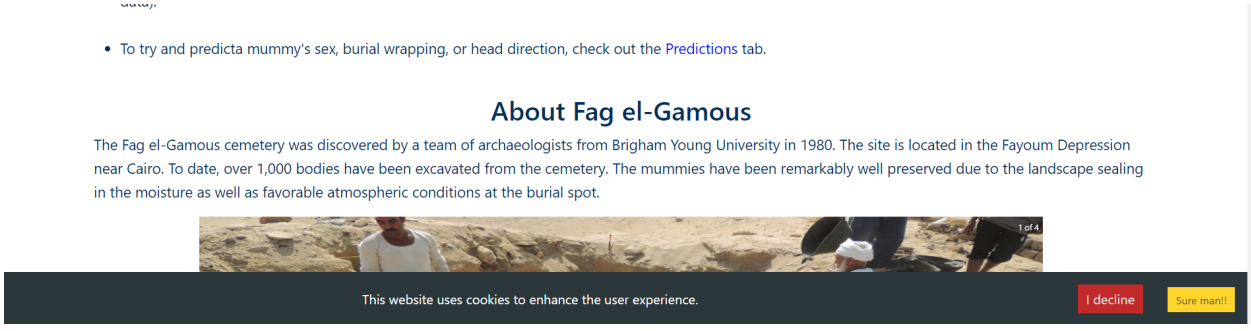
password and never sends it back to the client-side. We also have multiple clear local storage to clear the user credentials we stored for a short time. Below is our hashing code for the authentication.

```
SHA256 sha256 = SHA256.Create();
string pw = login.password;
pw = pw + pepper;
byte[] inputBytes = Encoding.UTF8.GetBytes(pw);
byte[] hashBytes = sha256.ComputeHash(inputBytes);

// Convert the byte array to a hexadecimal string
StringBuilder sb = new StringBuilder();
foreach (byte b in hashBytes)
{
   sb.Append(b.ToString("x2"));
}
string hashString = sb.ToString();
string code = "buffalo";
```

VI. Privacy

Privacy page is done and can be accessed by the footer that only appears after cookies are accepted.

Cookie Consent form works! Great!



VII. Other

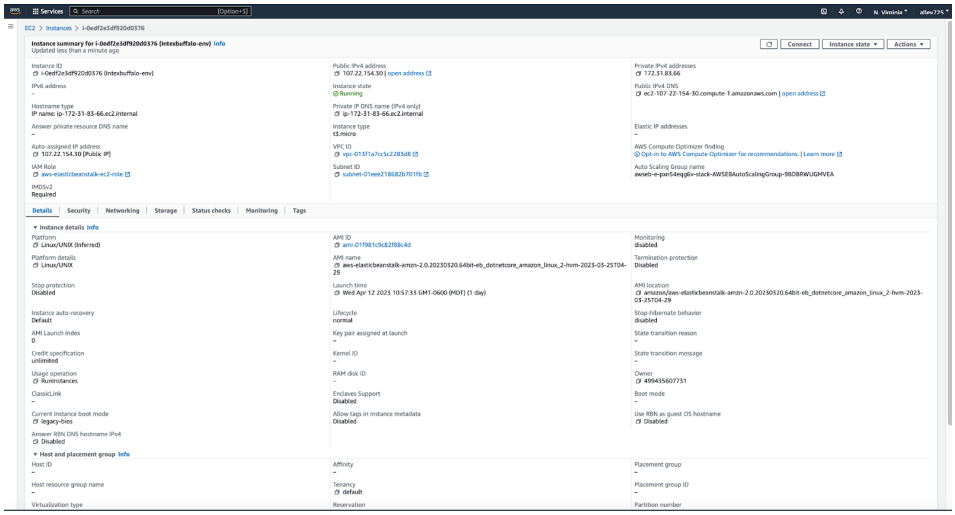
We create the our recommendations of Egypts 2020 data privacy law in the executive summary

AWS Services Used

We built and deployed our website using the following AWS Services:

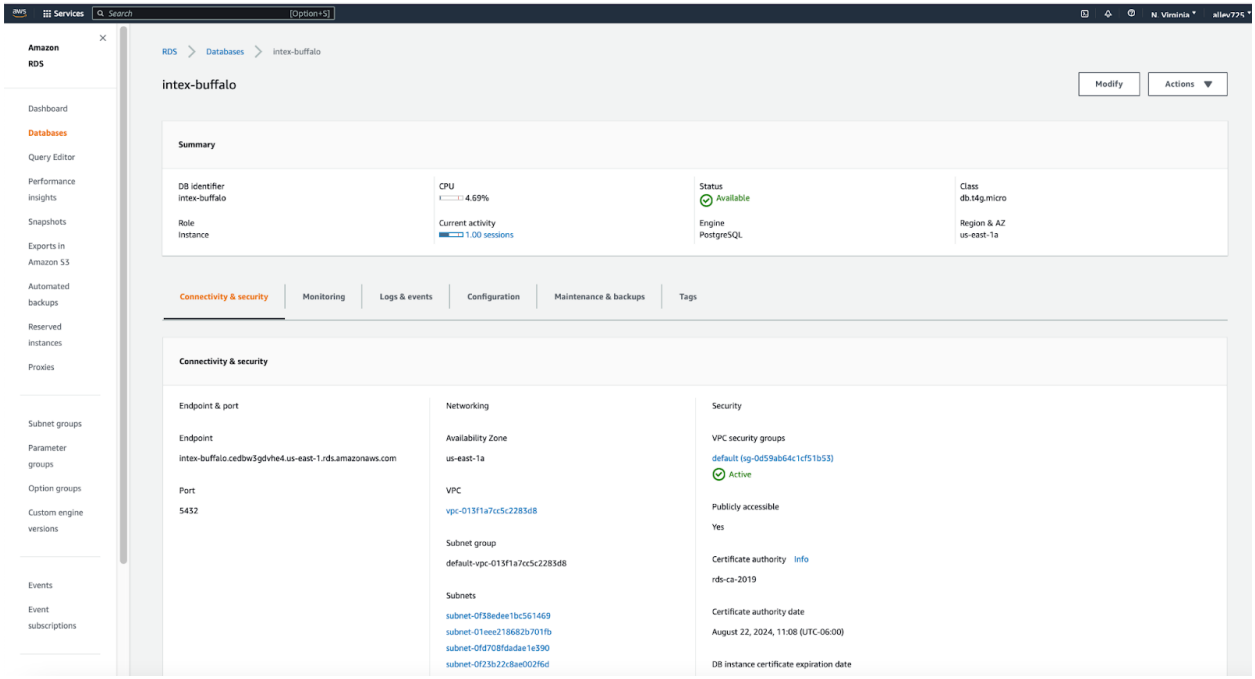
I. EC2

A. Used to host and server the website.



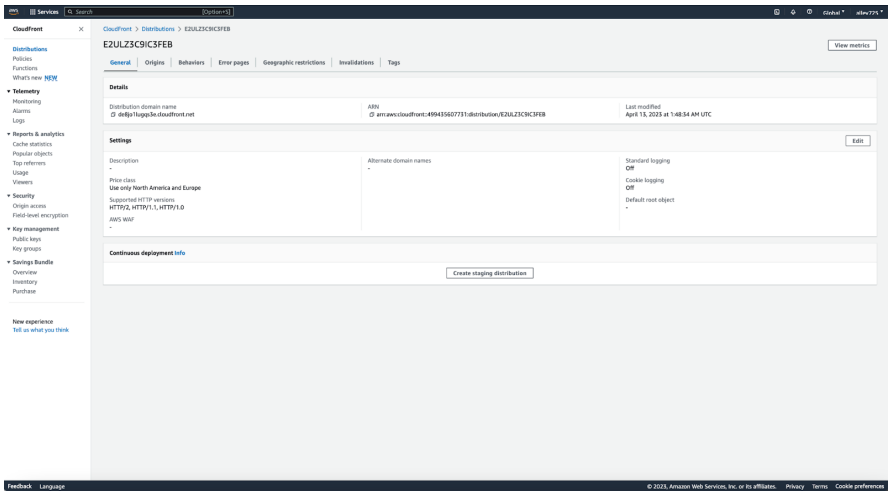
II. RDS

A. Used to host our PostgreSQL



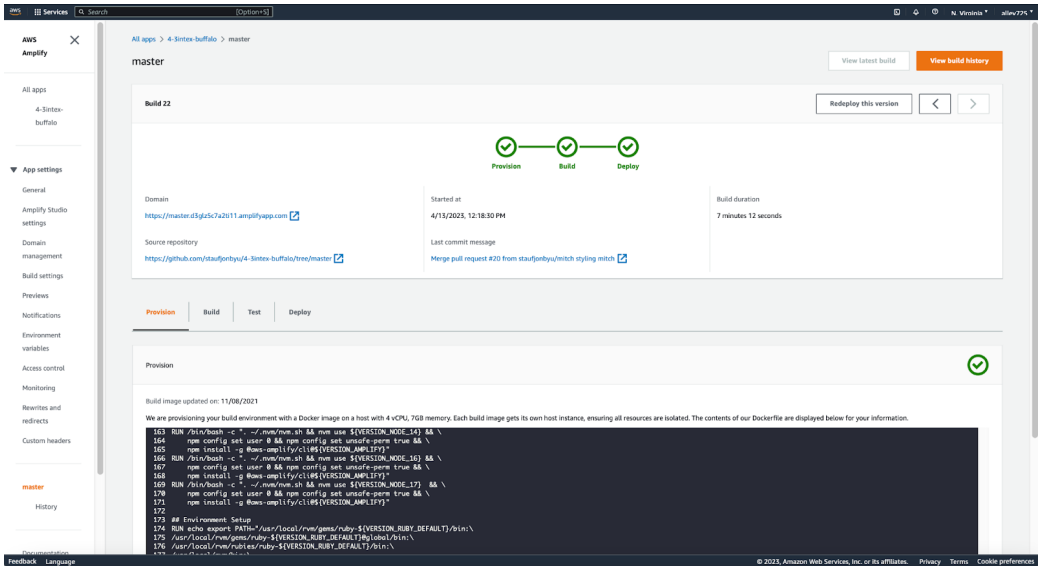
III. CloudFront

A. Used to redirect HTTP traffic to HTTPs traffic on Elastic Beanstalk



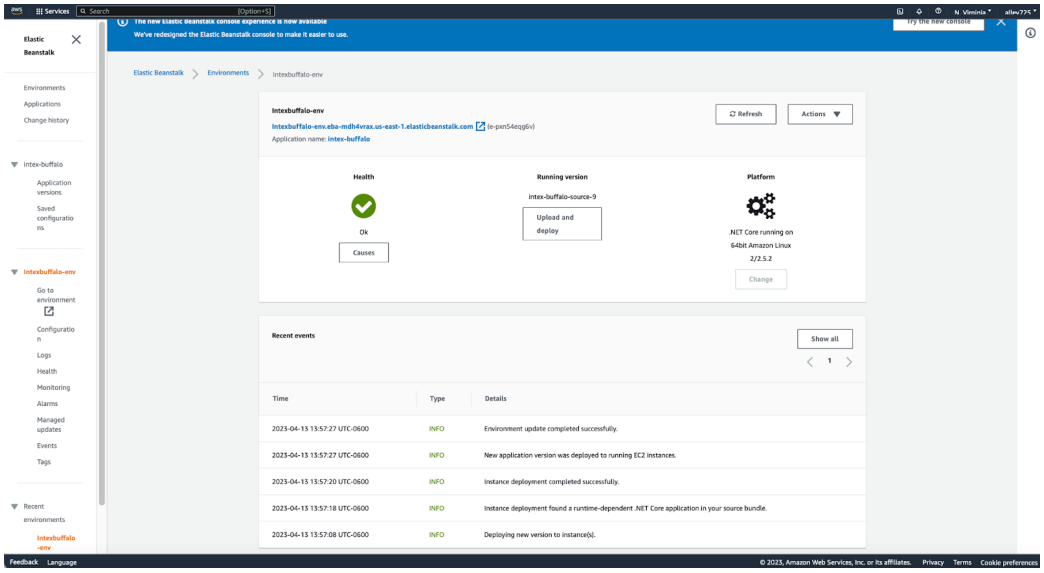
IV. Amplify

A. Used to support the frontend of the site



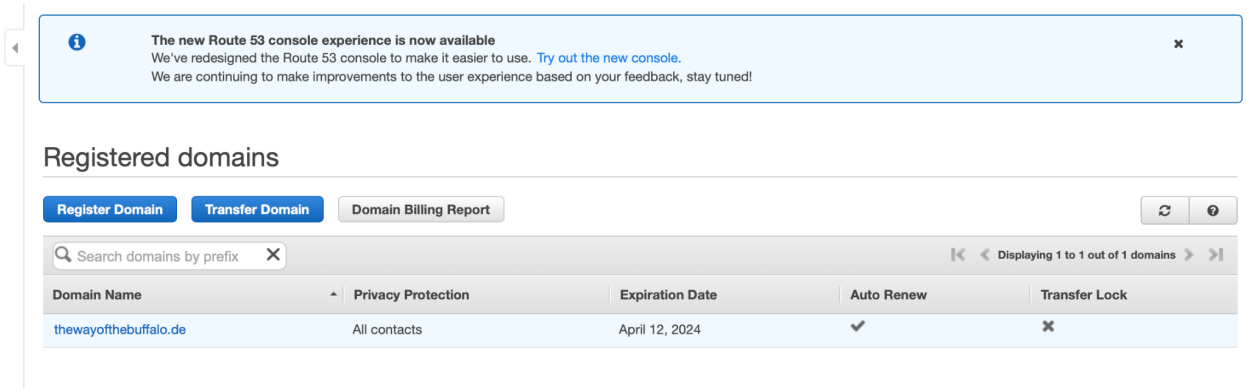
V. Elastic Beanstalk

A. Used to deploy and manage the site



VI. Route 53

A. Used to register our custom domain.



Recommended Services

AWS Lambda

One AWS service that would be valuable to implement in the future is Lambda. In Lambda, you can build and deploy machine learning models which can be called directly so that ONNX files will not need to be continually reloaded every time the model is trained.

Executive Summary

Introduction

The Fag el-Gamous cemetery, located on the eastern edge of the Fayoum Depression near Cairo, was discovered by a team of archaeologists from Brigham Young University in 1980. Since then, the remains of over 1,000 individuals have been exhumed from the site as well as numerous other artifacts.

For decades, archaeologists have been keeping records of the people and items uncovered at Fag el-Gamous. Currently, researchers study and measure the remains of bones to determine a mummy's sex, and other trends are identified by studying the data manually. But what if there were a better way?

Our team was tasked to design a website to help research teams better organize and utilize the data they collect from Fag el-Gamous. On the site, users can view burial records as well as filter them to find the information they are looking for. In addition, authorized users (administrators or researchers with credentials) have the ability to edit, create, and delete records from the database.

The primary functionality of our website is the utilization of machine learning models to make predictions about a burial based on a number of different factors. These models allow researchers, for example, to predict whether an exhumed body is male or female based on their approximate age at death, length, hair color, and how they were wrapped. The site also contains an "unsupervised" model, meaning it is not used to predict a specific label, but rather can be used to group together similar data points.

Basic Site Functionality

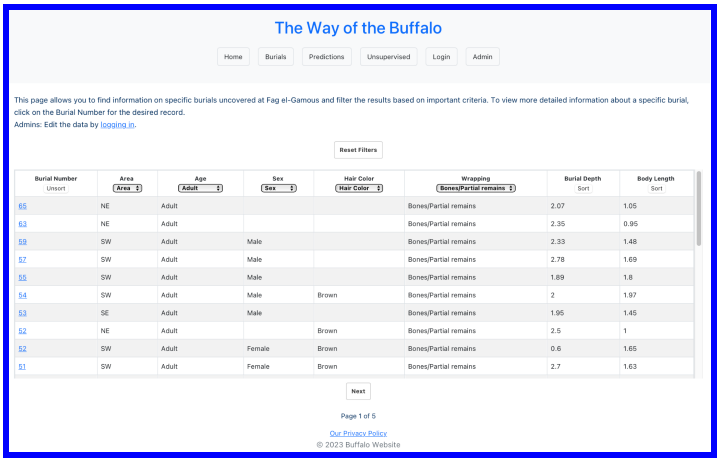
Home Page

Predictive/Analytical Models

- See “Models” below

Burial Information Table

- User can also view more detailed information on a particular record by clicking its burial number



Models

Data Cleaning

To get the data into a usable form after pulling it from our database, we needed to change several columns’ data types, drop columns that did not have a significant amount of

non-null values, and bin values to make them more standardized. In addition, we dropped any columns that did not contain information relevant to the findings from the burial.

We also needed to instantiate values in numerical columns because most of our models did not allow for null values. To do this, we used the column's mean value in place of any missing data.

Supervised (Predictive) Models

For our supervised models, we were asked to use burial information and textile data to predict the sex of the mummy. Due to limitations in the data, we were then asked to pivot and instead use wrapping state and head direction as our labels. In the end, we trained five supervised models:

1. Decision Tree Model to Predict Wrapping
2. Decision Tree Model to Predict Head Direction
3. Decision Tree Model to Predict Sex (using only Burial Main table data)
4. Decision Tree Model to Predict Sex (using Burial Main and Textile data)
5. Regression Model to Predict Wrapping

Feature Selection

When we first trained our decision tree classification models, we used all features that had not been taken out during the data cleaning process.

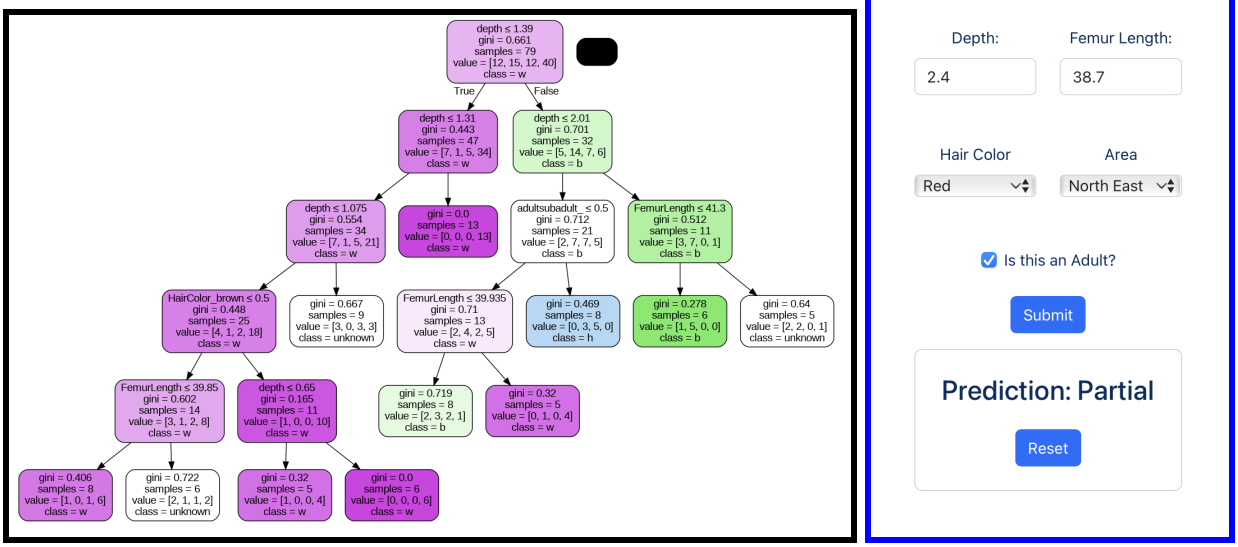
After training our model the first time, we ran feature importance calculations to determine how important each feature was in determining the label. We then filtered out all features that did not have any measurable impact.

We then re-trained the model using only the important features. This model is what we finally imported and deployed onto the website.

Model Visualization and Site Implementation

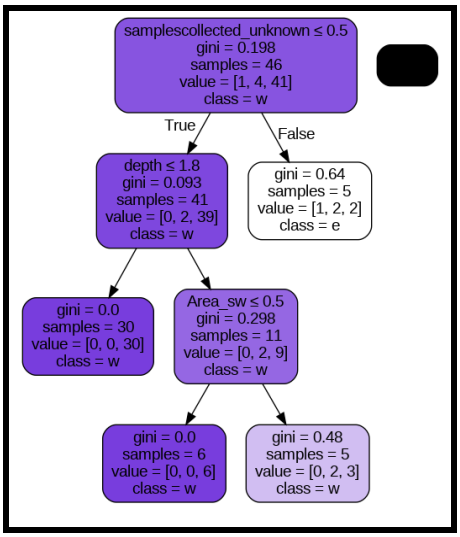
For decision tree classification models, the feature at the top of the tree is generally the most significant. If the condition in the node is met (True), the prediction follows the branch to the left. If not (False), the prediction follows the branch to the left.

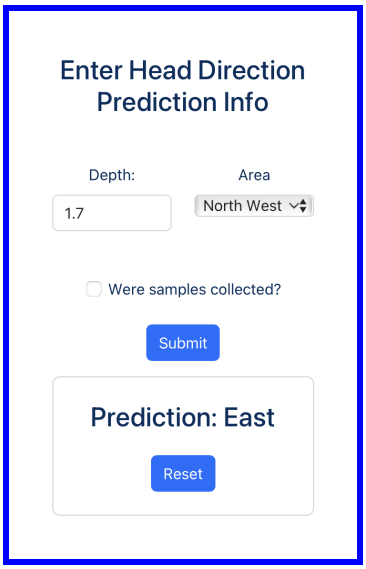
1. **Decision Tree Model to Predict Wrapping:**
    a. Accuracy Score: 0.56



2. Decision Tree Model to Predict Head Direction

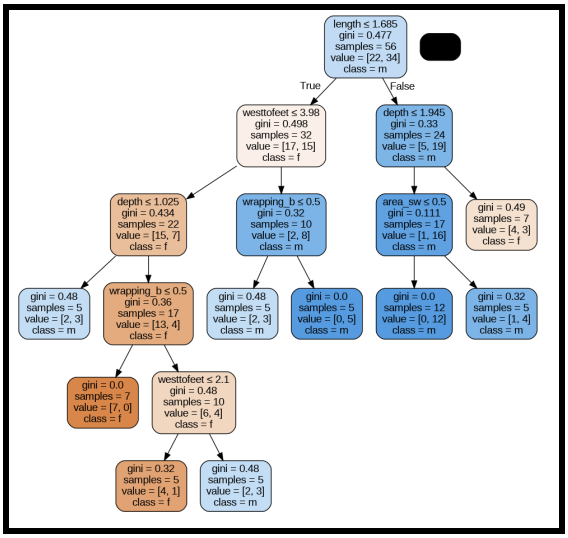
a. Accuracy Score: 0.86

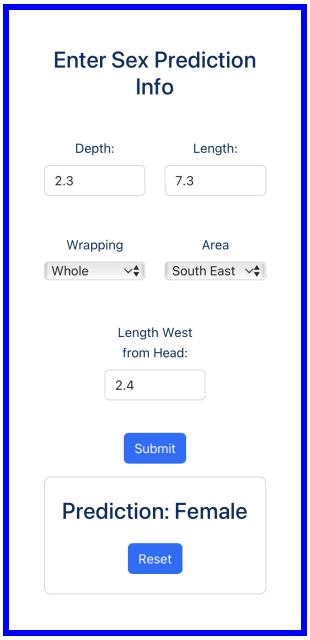




3. Decision Tree Model to Predict Sex (using only Burial Main table data)

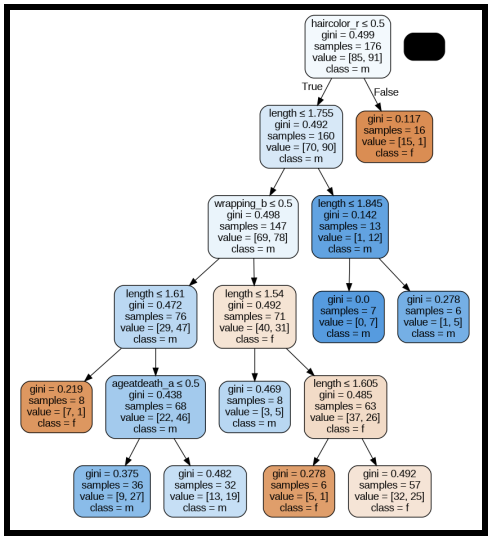
a. Accuracy Score: 0.48





4. Decision Tree Model to Predict Sex (using Burial Main and Textile data)

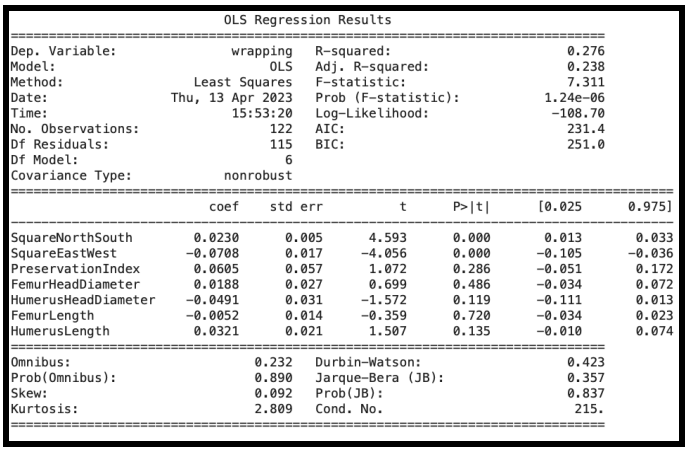
a. Accuracy Score: 0.71



5. Regression Model to Predict Wrapping

a. R-Squared: 0.276

b. This model is not deployed on our website.



OLS Regression Results

Dep. Variable:	wrapping	R-squared:	0.276
Model:	OLS	Adj. R-squared:	0.238
Method:	Least Squares	F-statistic:	7.311
Date:	Thu, 13 Apr 2023	Prob (F-statistic):	1.24e-06
Time:	15:53:20	Log-Likelihood:	-108.70
No. Observations:	122	AIC:	231.4
Df Residuals:	115	BIC:	251.0
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
SquareNorthSouth	0.0230	0.005	4.593	0.000	0.013	0.033
SquareEastWest	-0.0708	0.017	-4.056	0.000	-0.105	-0.036
PreservationIndex	0.0605	0.057	1.072	0.286	-0.051	0.172
FemurHeadDiameter	0.0188	0.027	0.699	0.486	-0.034	0.072
HumerusHeadDiameter	-0.0491	0.031	-1.572	0.119	-0.111	0.013
FemurLength	-0.0052	0.014	-0.359	0.720	-0.034	0.023
HumerusLength	0.0321	0.021	1.507	0.135	-0.010	0.074

Omnibus:	0.232	Durbin-Watson:	0.423
Prob(Omnibus):	0.890	Jarque-Bera (JB):	0.357
Skew:	0.092	Prob(JB):	0.837
Kurtosis:	2.809	Cond. No.	215.

Findings

In analyzing the models, the most significant takeaways come by considering which features played the biggest role in determining the labels. In the case of our models predicting wrapping and head direction, depth was a significant determining factor. This supports the idea that burial practices changed over time, since a deeper burial means that it is probably older than one that is shallower. For example, the decision tree indicates that older burials have less wrapping. This could mean that practices changed over time to wrap the bodies more.

In determining the sex of the bodies, length was a significant factor in both of our models. This has significant implications as it could give researchers and archaeologists a simpler way to determine (with varying degrees of accuracy) the sex without having to analyze detailed bone measurements.

One limitation of these models is the fact that their accuracy scores are not high enough to give definitive evidence that the predictions are correct. This deficiency likely stems from a lack of usable data that could be used to train the model. As researchers continue to use the site to record standardized, complete, and accurate data, the model will improve and its predictions will become more reliable.

Unsupervised Model – K-Modes

For the unsupervised portion, the K-modes clustering was chosen due to the significant numbers of categorical data that were used once the Burial Main table was connected with the other Textile tables.

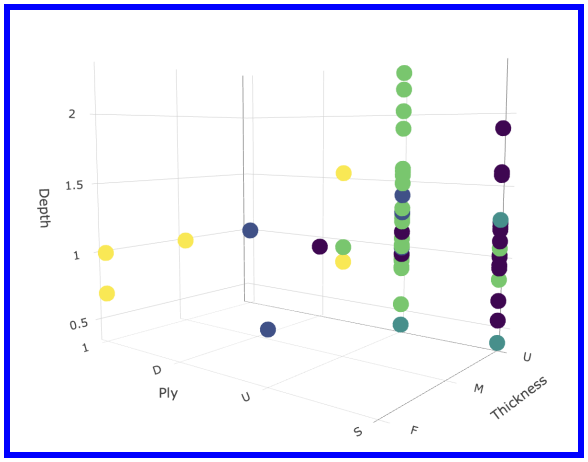
Feature Selection

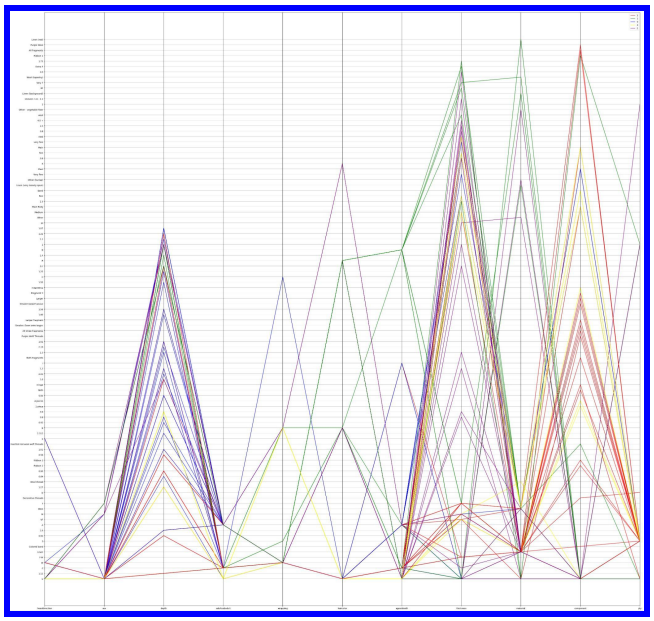
The columns that were chosen to run in the model were made by dropping all the IDs of the different tables that came when merged. Other columns were dropped based on best judgment on if it would be important in determining the relationship between textiles and burials. For example, the “excavationrecorder” was dropped because the recorder shouldn’t impact the textiles used for the burial.

Further columns were dropped as the model was run to decrease the number of columns that were used. For example, 'angle','preservation', and 'area' were dropped after looking through the visualizations and seeing that these columns distracted from the overall clustering. Also, recommended columns by the client were kept in.

The clustering number was determined as the K-Modes model was run. Because K-modes were used instead of K-means, the Elbow method and the Calinski-Harabasz (CH) Criterion method were not used. Based on recommendations, and seeing the proportions in each cluster, five clusters were chosen. This amount gave a somewhat distributed clustering model.

Model Visualization and Site Implementation





Findings

The three main findings had to do with the Sex, Wrapping Material, and Head Direction.

- We see that as the burials got deeper that Linen Textiles was used more often than Wool Textiles. In the depth of 0-0.5, Wool is used about 75% of the time and Linen about 25%. In depth of 0.5-2, it flips. Linen is 75% and Wool is 25%.
- In further analysis, we also see that Females that were found in burials were the ones found with more colorful and decorative materials. In the clusters with mostly females, like 1 and 2, we see materials like Purple Weft Threads, Decorative Thread, and Ribbons.
- As we looked further into the specific direction of Head Direction, we noticed that the most common direction was West. But with Cluster 2, the main head direction we see is East. This cluster has Brown and Red Hair Colors, had a varied depth, and had a majority of Females. This could possibly mean that women with colored hair might have had the majority of East Head Direction.

Cost Estimate

Assuming the site will get ~100 monthly users and all costs will be paid monthly, we estimate that our site will cost approximately $293.18 per month to run on AWS. This estimate accounts for $5.33 for a single t4g.small EC2 instance and $287.85 to run our database on Amazon RDS.

Additional costs will be incurred from the development hours necessary to maintain the site, but these should be minimal as the website is not very CPU-intensive and should not require too much upkeep. We estimate approximately $200 for development costs.

Security Recommendations

As it stands, the website is GDPR compliant and abides by generally-accepted cybersecurity standards. That being said, it should be noted that Egypt passed a data privacy law in 2020 that sets requirements beyond that of GDPR and may affect the site if it deals with data in Egypt.

This legislation (Law No. 151) prohibits the processing of any personal data without explicit consent from the individual. It also gives individuals numerous rights such as restricting access to data, revoking prior consent, and being promptly informed of any data breach. Site admins should be aware of this law and where compliance may be necessary.

Conclusion

The website we have developed will help archaeologists and researchers now and in the future to record, analyze, and predict data related to uncovered burials at Fag el-Gamous. As more data is collected, the models can be re-trained and become even more accurate with time.