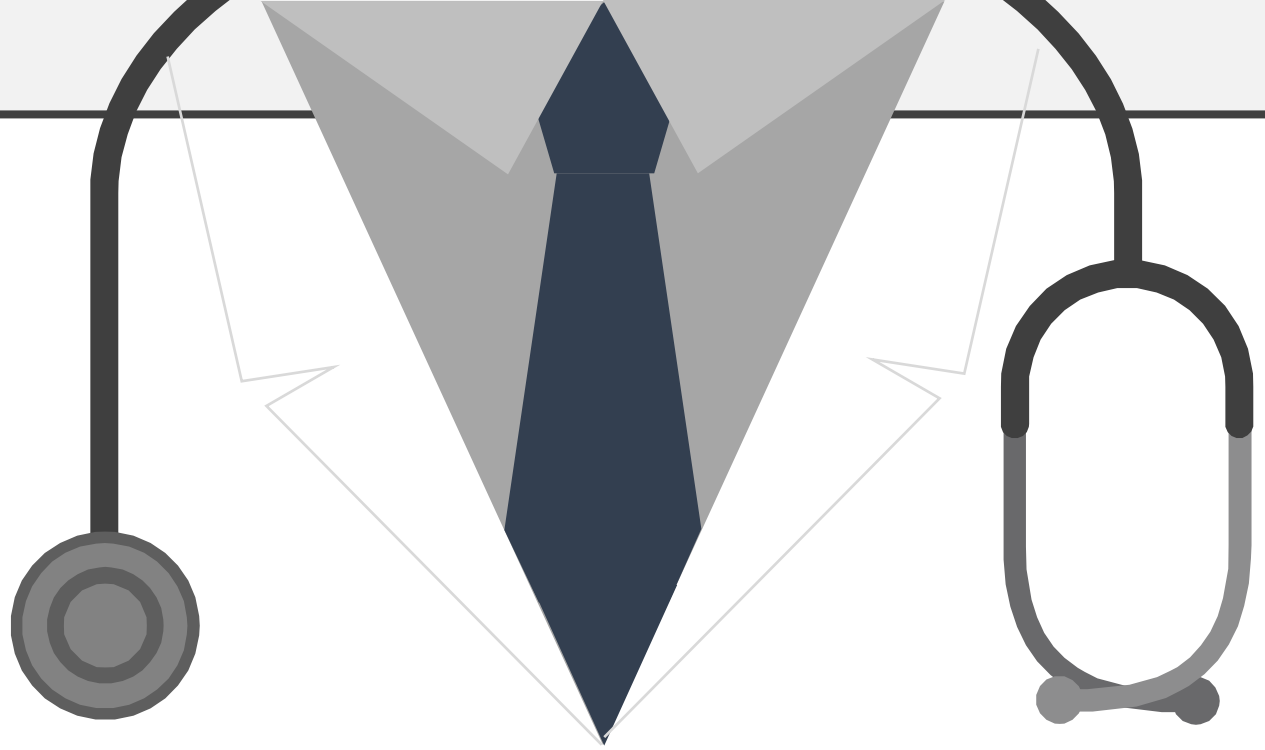
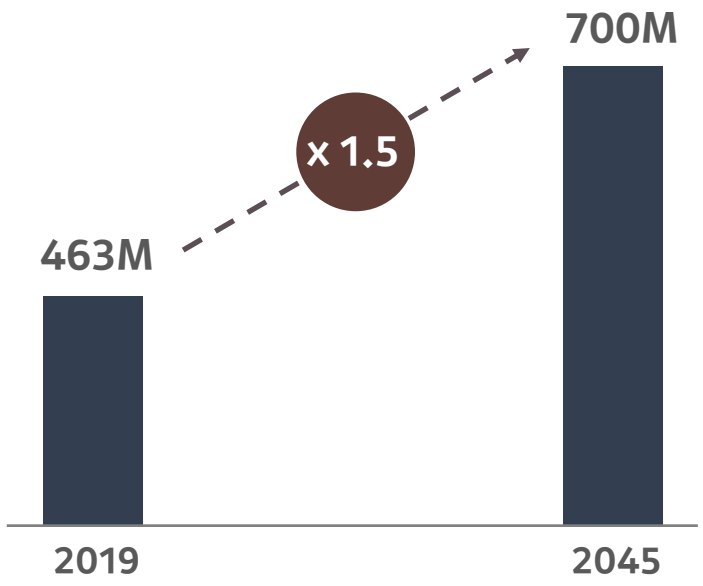


*PREDICTING **TYPE II DIABETES** WITH EARLY SYMPTOMS*

IDS 702 FINAL PROJECT | RHAYOUNG PARK

BACKGROUND

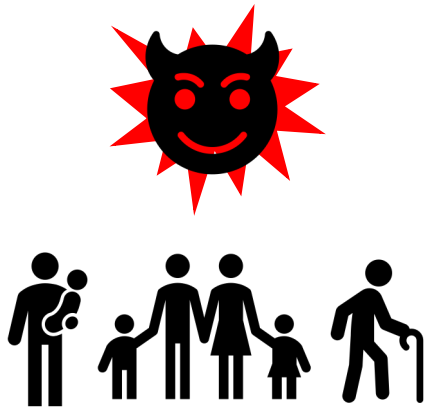
INCREASING DIABETES PATIENTS



1.7% of Total GDP

Diabetes related healthcare cost

7th leading cause of death



80,000 death per year

Reference: American Diabetes Association

DATA

COLLECTION

- Asked direct questions to Sylhet Diabetes Hospital patients
- Total 520 observations with 16 predictors and binary response variable
- No missing values

VARIABLES

- Continuous
- Binary
- Response

- Age
- Sex
- Polyuria / Polydipsia / Polyphagia
- Sudden weight loss / Weakness
- Genital thrush
- Visual blurring
- Itching / Irritability
- Delayed healing
- Partial paresis / Muscle stiffness
- Alopecia
- Obesity

- Diabetes

QUESTIONS OF INTEREST

SIGNIFICANT EARLY SYMPTOMS

FOCUSING ON ‘ITCHING’

50%

INTERACTIONS RELATED TO ‘ITCHING’

EXPLORATORY DATA ANALYSIS

Itching and diabetes appears to be not related

	Has developed diabetes	Has not developed diabetes
Has itching	48.1%	49.5%
Does not have itching	51.9%	50.5%

chi-squared test : 0.83

Insufficient data for some age categories

	Below 31	Age 31 - 40	Age 41 - 50	Age 51 - 60	Age 61 - 70	Over 71
Has developed diabetes	15	84	87	78	49	7
Has not developed diabetes	30	87	58	49	17	7

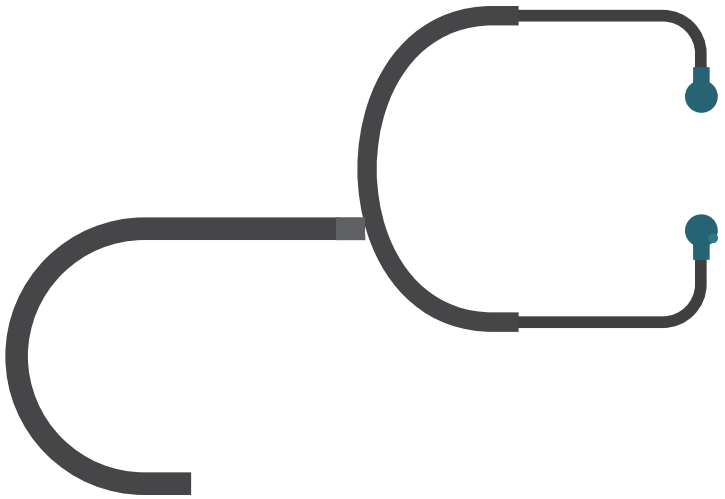
FINAL MODEL: INTRODUCTION

FINAL MODEL

$$
\begin{aligned}
logit\hat{y} = {} & \hat{\beta}_0 \\
& + \hat{\beta}_d x_{gender} \\
& + \hat{\beta}_a x_{itching} \\
& + \hat{\beta}_b x_{polydipsia} + \hat{\beta}_c x_{polyuria} + \hat{\beta}_h x_{polyphagia} \\
& + \hat{\beta}_e x_{irritability} \\
& + \hat{\beta}_f x_{genitalthrush} \\
& + \hat{\beta}_g x_{partialparesis} \\
& + \hat{\beta}_j x_{visualblurring} \\
& + \hat{\beta}_i x_{agecentered} \\
& + \hat{\beta}_k x_{itching:visualblurring}
\end{aligned}
$$

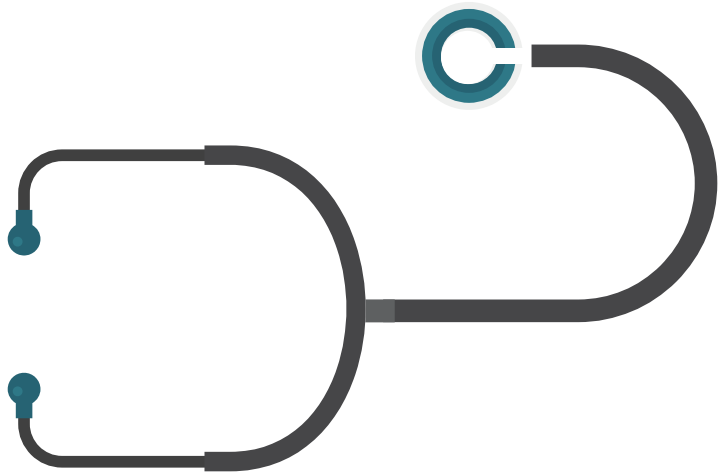
	Coefficient	P-Value
Intercept	0.24	0.64
Male	-4.73	4.5e-14 ●●●
Has itching	-1.75	0.01 ●
Has polydispsia	- 5.42	2.2e-12 ●●●
Has polyuria	4.29	1.6e-11 ●●●
Has polyphagia	1.23	0.01 ●
Has irritability	2.66	2.6e-0.5 ●●●
Has genital thrush	1.73	0.002 ●●
Has partial paresis	1.15	0.02 ●
Has visual blurring	2.83	0.003 ●●
Age (centered)	-0.07	0.002 ●●
Has itching and visual blurring	-3.01	0.006 ●●

FINAL MODEL: INTERPRETATION

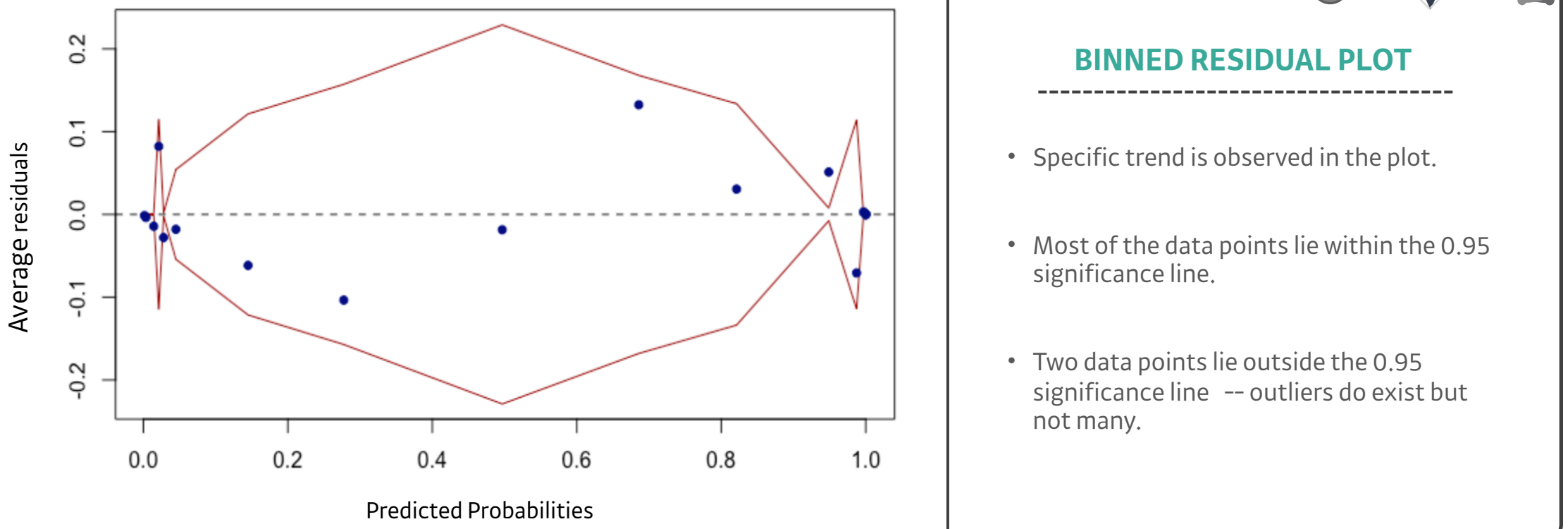


- All predictors except the intercept was statistically significant at 0.95 level.
- Holding everything else constant, the odds of developing diabetes were **83% less** if the person experienced itching, compared to people who did not experience itching.

- Interaction between visual blurring and itching was significant. The odds of developing diabetes were **95% lower** for people who had both visual blurring and itching, holding everything else constant.
- Polydipsia, polyuria and polyphagia were positively associated with diabetes, which coincides with earlier studies.



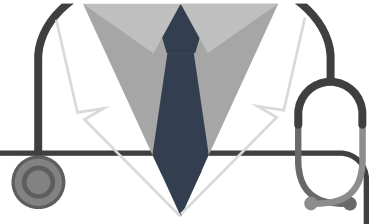
FINAL MODEL: ASSESSMENT (1/3)



BINNED RESIDUAL PLOT

- Specific trend is observed in the plot.
- Most of the data points lie within the 0.95 significance line.
- Two data points lie outside the 0.95 significance line -- outliers do exist but not many.

FINAL MODEL: ASSESSMENT (2/3)



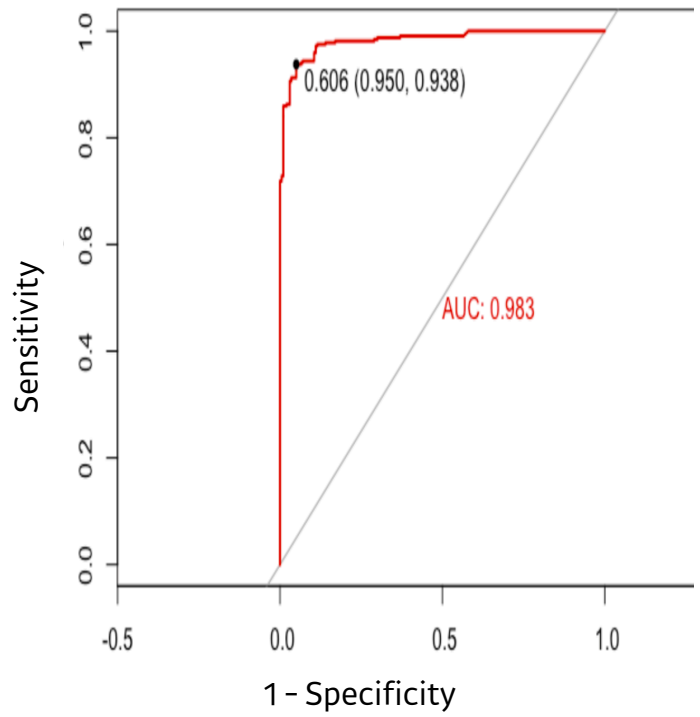
	VIF		VIF
Itching	3.16	**Has genital thrush**	1.61
Male	2.10	**Has partial paresis**	1.24
Has polydipsia	1.92	**Age (centered)**	1.99
Has polyuria	2.07	**Has visual blurring**	5.81
Has polyphagia	1.44	**Has itching and visual blurring**	6.84
Has irritability	1.42		

MULTICOLLINEARITY – VIF

- Age was centered to reduce multicollinearity. As a result, the VIF values were low.
- Two predictors had notable VIF values (visual blurring and the interaction term). These were kept in the model as it was not larger than 10.
- Interactions that include categorical variables are correlated in nature and thus it might have relatively high VIF scores compared to other predictors.

FINAL MODEL: ASSESSMENT (3/3)

ROC CURVE



0.98

AUC VALUE

- This model predicts diabetes correctly 98% of the time

CONFUSION MATRIX

	Value
Accuracy	**0.92**
Specificity	**0.90**
Sensitivity	**0.94**

94% of the time, the model will correctly predict people who will have diabetes,
but 6% of the time, it generates false-negatives

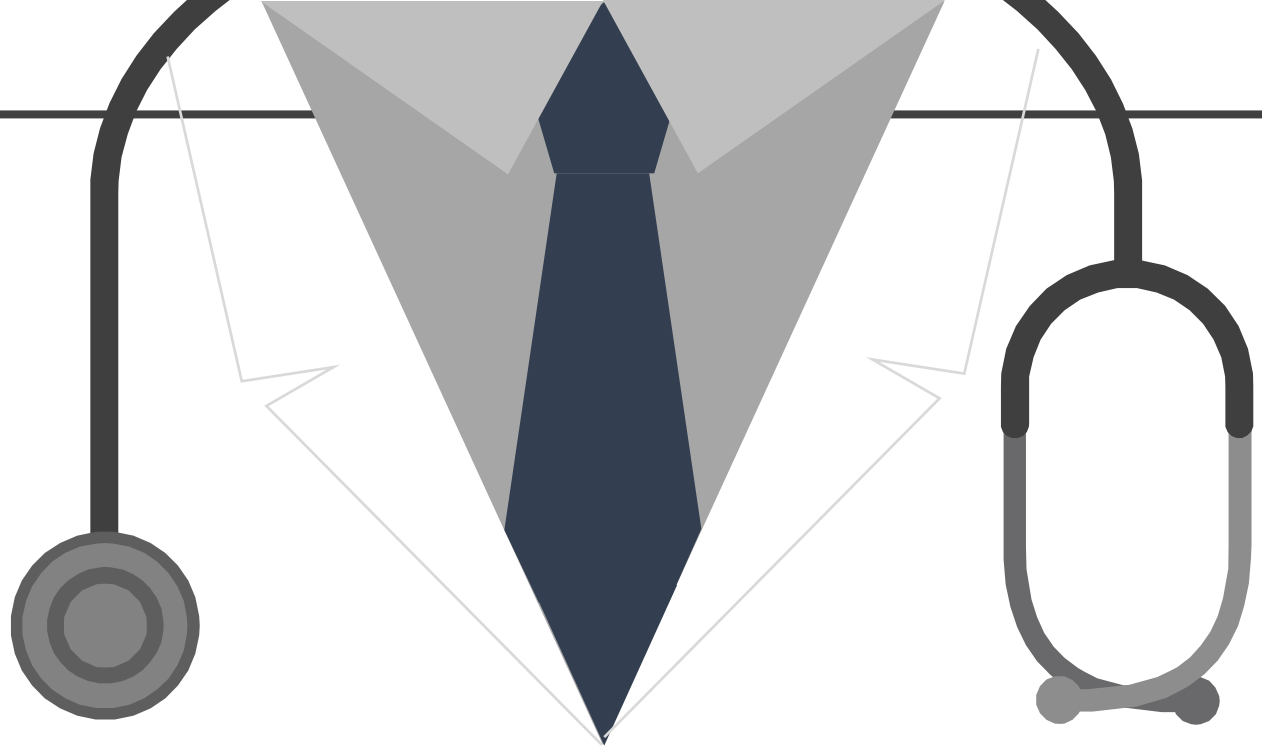
CONCLUSION

LIMITATION

- Data points in binned residual plots were not randomly distributed.
- Some potentially significant predictors, such as physical activity, BMI index and socioeconomic status were not provided in the dataset.

SUMMARY

- Significant predictor of diabetes were itching, sex, polydipsia, polyuria, polyphagia, irritability, genital thrush, partial paresis, visual blurring, age and interaction between itching and visual blurring
- Itching and interaction between itching and visual blurring were negatively associated with developing diabetes.



THANK YOU !