Question 4 (Variant 3)[25 marks]

(a) **Binary Classification (6 Marks)**
For following binary classification outcome table, calculate the following appraisal metrics.

  (i) (1 Mark) accuracy;

 (ii) (1 Mark) recall;

(iii) (1 Mark) precision;

(iv) (1 Mark) F-measure.

	Predict Negative	Predict Positive
Observed Negative	9530	10
Observed Positive	300	160

  (v) (2 Marks) Explain why the F-measure is considered a more informative measure of performance than the Accuracy score.

(b) **Inference Procedures (10 Marks)**
Two IT training companies, *XtraTech* and *YourSkills*, offer an exam preparation course for a well-known computer industry certification. A study was carried out to compare the results from the most recent group of students from both companies.

- 30 students from the *XtraTech* course have completed the test. The average score for these students was 910 marks with a standard deviation of 48 marks.

- 25 students from the *YourSkills* course have completed the test. Their average score was 950 marks with a standard deviation of 42 marks.

Test the hypothesis that the both sets of students perform equally well on average. You may use a significance level of 5%. You may assume that both samples are normally distributed and have equal variance.

  (i) (2 Marks) Formally state the null and alternative hypotheses for this procedure.

 (ii) (2 Marks) Compute the point estimate for the difference in means of the results from both courses.

(iii) (2 Marks) Compute the appropriate value for standard error for this test. Clearly show your workings.

(iv) (2 Marks) Compute the test statistic.

 (v) (2 Marks) What is your conclusion for this procedure?

(c) **Inference Procedures (9 Marks)**

A study finds that 45% of IT users out of a random sample of 500 in a large community preferred one web browser to all others. In another large community, 30% of IT users out of a random sample of 390 prefer the same web browser.

(i) (2 Marks) Compute the point estimate for the difference in proportions of IT users who prefer this particular web browser.

(ii) (4 Marks) Compute a 95% confidence interval for this difference in proportions.

(iii) (3 Marks) Based on this confidence interval, test the hypothesis that the proportion of IT users using this web browser is the same for both communities. State your null and alternative hypotheses clearly.